

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 20, 2001, 23:48:51 ; Search time 2434.93 Seconds  
(without alignments)  
3620.891 Million cell updates/sec

Title: US-09-558-474-1  
Perfect score: 570  
Sequence: 1 ATGCTGGGAGCAGAGCTGT.....CAGCAACCTGAGTCCCTAA 570

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl :  
1: gb\_ba1 :  
2: gb\_ba2 :  
3: gb\_ba3 :  
4: gb\_in1 :  
5: gb\_in2 :  
6: gb\_in3 :  
7: gb\_om :  
8: gb\_ov :  
9: gb\_pat1 :  
10: gb\_pat2 :  
11: gb\_ph :  
12: gb\_p11 :  
13: gb\_p12 :  
14: gb\_p13 :  
15: gb\_p14 :  
16: gb\_ba1 :  
17: gb\_ba2 :  
18: gb\_fun :  
19: em\_htgo\_hum :  
20: em\_htgo\_inv :  
21: em\_htgo\_rod :  
22: em\_htg\_hum1 :  
23: em\_htg\_hum2 :  
24: em\_htg\_hum3 :  
25: em\_htg\_hum4 :  
26: em\_htg\_hum5 :  
27: em\_htg\_hum6 :  
28: em\_htg\_hum7 :  
29: em\_htg\_hum8 :  
30: em\_htg\_inv1 :  
31: em\_htg\_inv2 :  
32: em\_htg\_other :  
33: em\_htg\_rod :  
34: em\_hum1 :  
35: em\_hum2 :  
36: em\_hum3 :  
37: em\_hum4 :  
38: em\_hum5 :  
39: em\_hum6 :  
40: em\_hum7 :  
41: em\_in :  
42: em\_om :  
43: em\_or :

44: em\_ov :  
45: em\_pat :  
46: em\_ph :  
47: em\_pl :  
48: em\_ro :  
49: em\_sts :  
50: em\_sy :  
51: em\_un :  
52: em\_vi :  
53: gb\_sts1 :  
54: gb\_sts2 :  
55: gb\_sts3 :  
56: gb\_sy :  
57: gb\_un :  
58: gb\_vil :  
59: gb\_vil2 :  
60: gb\_vil3 :  
61: gb\_vil4 :  
62: gb\_vil5 :  
63: gb\_vil6 :  
64: gb\_vil7 :  
65: gb\_vil8 :  
66: gb\_vil9 :  
67: gb\_vil10 :  
68: gb\_vil11 :  
69: gb\_vil12 :  
70: gb\_vil13 :  
71: gb\_vil14 :  
72: gb\_vil15 :  
73: gb\_vil16 :  
74: gb\_vil17 :  
75: gb\_vil18 :  
76: gb\_vil19 :  
77: gb\_vil20 :  
78: gb\_vil21 :  
79: gb\_vil22 :  
80: gb\_vil23 :  
81: gb\_vil24 :  
82: gb\_vil25 :  
83: gb\_vil26 :  
84: gb\_vil27 :  
85: gb\_vil28 :  
86: gb\_vil29 :  
87: gb\_vil30 :  
88: gb\_vil31 :  
89: gb\_vil32 :  
90: gb\_vil33 :  
91: gb\_vil34 :  
92: gb\_vil35 :  
93: gb\_vil36 :  
94: gb\_vil37 :  
95: gb\_vil38 :  
96: gb\_vil39 :  
97: gb\_vil40 :  
98: em\_ba3 :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description        |
|------------|-------|---------------|--------|----|--------------------|
| 1          | 570   | 100.0         | 570    | 10 | AX097621 Sequence  |
| 2          | 570   | 100.0         | 1025   | 89 | AF301620 Homo sapi |
| 3          | 566.8 | 99.4          | 1026   | 85 | AB030000 Homo sapi |
| 4          | 566.8 | 99.4          | 1055   | 9  | AX048200 Sequence  |
| 5          | 391   | 68.6          | 1203   | 10 | AX097623 Sequence  |
| 6          | 391   | 68.6          | 1358   | 94 | AF301619 Mus muscu |
| 7          | 192.8 | 33.8          | 4039   | 85 | AB030001 Homo sapi |
| 8          | 192.8 | 33.8          | 155929 | 69 | AC025574 Homo sapi |









```

Db 353 CAAGGACTCAAGGACACAGCCAGTTCTGCTTGCAAGAGATCCGCCAAGGCTGGCTTTT 412
Qy 238 TATGAGAGCTGCTAGGATCGGATATTTTTCACAGGGAGCCTTCTGCTCCCTGATAGC 357
Db 413 TATAAGCACCTGCTGACTCTGACATCTTCAAGGGAGCCTGCTACTCCCTGATAGC 472
Qy 358 CTTGTGGCGAGCTTCATGCTCCCTACTGCGCCTTCAGCCAACTCCCTCAGCCTGAGGCT 417
Db 473 CCCATGGAGCAACTTCACACCTCCCTACTAGACTCAGCCAACTCCCTCCAGCAGAGAT 532
Qy 418 CACCAGTGGAGACTCAGCAGATATCCAAAGCTCAGTCCAGCCAGCAGCAGCAGCTC 477
Db 533 CACCCCGGAGAGACCAACAGATGCCAGCTGAGTTCTAGTACAGCAGTGGCAGCGCCC 592
Qy 478 CTTCTCCGCTTCAAACTCTTCAGAGCTCCAGGCTTTTGGTGTAGCGCCCGGCTC 537
Db 593 CTTCTCCGCTTCAAACTCTTCAGAGCTCCAGGCTTTTGGTGTAGCGCCCGGCTC 652
Qy 538 TTTGCCATGGAGCAGCAGCAACCTGAGT 564
Db 653 TTTGCCATGGAGCAGCAGCAACCTGAGT 679

RESULT 6
AF301619 AF301619 1358 bp mRNA ROD 04-DEC-2000
LOCUS Mus musculus interleukin 23 p19 subunit mRNA, complete cds.
ACCESSION AF301619
VERSION AF301619.1 GI:11528340
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1358)
AUTHORS Oppmann, B., Lesley, R., Blom, B., Timans, J.C., Xu, Y., Hunte, B.,
Vega, F., Yu, N., Wang, J., Singh, K., Zonin, F., Vaisberg, E.,
Churakova, T., Liu, M., Gorman, D., Wagner, J., Zurawski, S., Liu, Y.,
Abrams, J.S., Moore, K.W., Rennick, D., de Waal-Malefyt, R., Hannum, C.,
Bazan, J.F. and Kastelein, R.A.
TITLE Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with
biological activities similar as well as distinct from IL-12
JOURNAL Immunity 13 (5), 715-725 (2000)
MEDLINE 20567322
REFERENCE 2 (bases 1 to 1358)
AUTHORS Kastelein, R.A., Gorman, D., Timans, J.C., Oppmann, B. and Bazan, J.F.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-2000) Molecular Biology, DNAX Research Institute,
901 California Avenue, Palo Alto, CA 94304, USA
FEATURES
source
1. .1358
/organism="Mus musculus"
/db_xref="taxon:10090"
113. .703
/note="four helical bundle cytokine; combines with the p40
subunit of IL-12 to form a novel, biologically active,
composite cytokine, IL-23"
/codon_start=1
/product="interleukin 23 p19 subunit"
/protein_id="AAG37231.1"
/db_xref="GI:11528341"
/translation="MLDCRAVIMLLPWVTOGLAVPRSSPDWAQCOQLSRNLCMLA
WNAHAPAGHNNLREDEETKNNVPRIQEDGCDGDLKDNDFCLORIQGLAFYK
HLSDSDFKGEPLLPSPEQLHTSLGLSQLLQPEDHPRETOQMFLSSSQWQRP
LLRSKILRSQAFALAIARVFAHGAATLEPLVPTA"
BASE COUNT 383 a 326 c 338 g 311 t
ORIGIN

Query Match 68.6%; Score 391; DB 94; Length 1358;
Best Local Similarity 81.8%; Pred. No. 3e-101;
Matches 464; Conservative 0; Mismatches 100; Indels 3; Gaps 1;

```

```

Qy 1 ATGCTGGGAGCAGAGCTGTAATGCTGCTGTGCTGCTGCTGACAGCTCAGGGCAGA 60
Db 113 ATGCTGGATTCAGAGCAGTAATAATGCTATGGCTGTTGGCCCTGGTCACTCAGGGCCTG 172
Qy 61 GCTGTGCTGGGGCAGCAGCCCTGCTGAGCTCAGTGGCCAGCAGCTTTCACAGAAGCTC 120
Db 173 GCTGTGCTAGGAGTAGCAGTCTGACTGGCTCAGTGGCCAGCAGCTCTCTCGGAATCTC 232
Qy 121 TGCACACTGGCTGGAGTGACATCCACTAGTGGACACATGGAT--CTAAGAGAAGAG 177
Db 233 TGCATGTCTAGCTTGAAGCGCACATGCCAGCGGACATATGAATCTCTAAGAGAAGAA 292
Qy 178 GGAGATGAAGAGACTTACAAATGATGTTCCCATATCCAGTGTGGAGATGGCTGTACACCC 237
Db 293 GAGATGAAGAGACTTAAATAATGTTGCCCGTATCCAGTGTGAAGATGGTGTGACCCA 352
Qy 238 CAAGAGACTCAGGGACACAGTTCAGTTCGTGTGCAAGAGATCCACAGGGTCTGATTTT 297
Db 353 CAAGGACTCAAGAGACACAGCAGTCTGCTTGTGAAGGATCCGCCAAGGTCTGGCTTTT 412
Qy 298 TATGAGAGCTGCTAGGATCGATATTTTCACAGGGGAGCCTTCTCTGCTCCCTGATAGC 357
Db 413 TATAAGCACCTGCTTGACTTGACATCTTCAAGGGGAGCCTGCTCTCTCCCTGATAGC 472
Qy 358 CTTGTGGCGCAGCTTTCATGCTCCCTACTGCGCCTCAGCCAACTCTCTCAGCCTGAGGCT 417
Db 473 CCCATGGAGCAACTTCACACCTCCCTACTAGACTCAGCCAACTCTCTCCAGCAGAGAT 532
Qy 418 CACCAGTGGAGACTCAGCAGATATCCAAAGCTCAGTCCAGCCAGCAGCTGAGCAGCTC 477
Db 533 CACCCCGGAGAGACCAACAGATGCCAGCCTGAGTTCTAGTACAGCAGTGGCAGCGCCC 592
Qy 478 CTTCTCCGCTTCAAACTCTTCAGAGCTCCAGGCTCCAGGCTTTGCGGTGAGCCCGGCTC 537
Db 593 CTTCTCCGCTTCAAACTCTTCAGAGCTCCAGGCTTTTGGGATAGTGTGCGCGGCTC 652
Qy 538 TTTGCCATGGAGCAGCAGCAACCTGAGT 564
Db 653 TTTGCCATGGAGCAGCAGCAACCTGAGT 679

RESULT 7
AB030001 AB030001 4039 bp DNA PRI 04-APR-2000
LOCUS Homo sapiens gene for SGFR, complete cds.
ACCESSION AB030001
VERSION AB030001.1 GI:7416074
KEYWORDS SGFR.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4039)
AUTHORS Hirata, Y. and Kosuge, Y.
TITLE SGFR; a novel member of the IL-6/G-CSF family
JOURNAL Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 4039)
AUTHORS Hirata, Y. and Kosuge, Y.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-1999) to the DBJ/EMBL/GenBank databases. Yuichi
Hirata, Chugai Research Institute for Molecular Medicine, Gene
Search Program; 153-2 Nagai, Nihari-mura, Ibaraki 300-4101, Japan
(E-mail: hiratayu@ciimmed.com, Tel:81-298-30-6211(ex.288),
Fax:81-298-30-6270)
FEATURES
source
1. .4039
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q13"
38. .43
318. .622
/product="SGFR"
TATA_signal
exon

```

```

• gene      join(461..622,842..940,1107..1253,1359..1520)
             /gene="SGRF"
CDS         join(461..622,842..940,1107..1253,1359..1520)
             /gene="SGRF"
             /codon_start=1
             /product="SGRF"
             /protein_id="BA093687.1"
             /db_xref="GI:7416075"
             /translation="MLGSRAVMLLLPMTAOGRAVPGSSPAWTCQOOLSKLCTLA
WAHPLVGHMDLREGDETTNDVPHIQCGDCDPOGLRDNDSQFCLORHQGLIYEX
LGSDFTEGPELSPDPSVQGLHSLGLQPEGHWHETQQLPSUSPQWQRLLE
LRFKILRSLOAEVAVARFAHAATLSP"
             842..940
exon        /gene="SGRF"
             /product="SGRF"
             1107..1253
exon        /gene="SGRF"
             /product="SGRF"
             1359..1826
             /product="SGRF"
             1803..1808
polyA_signal 1098 a 940 c 1029 g 972 t
ORIGIN
Query Match      33.8%; Score 192.8; DB 85; Length 4039;
Best Local Similarity 74.3%; Pred. No. 1.9e-44;
Matches 309; Conservative 0; Mismatches 2; Indels 105; Gaps 1;

QY 260 AGTCTGCTGCAAGAGATCCACCAGGGTCTGATTTTATGAGAGCTGCTAGGATCGG 319
      |||||||
Db 1105 AGTCTGCTGCAAGAGATCCACCAGGGTCTGATTTTATGAGAGCTGCTAGGATCGG 1164
      |||||||

QY 320 ATATTTTCAGGGAGCGCTTCTCTCCTCGATAGCCCTGTGGCGAGCTTCATGCCT 379
      |||||||
Db 1165 ATATTTTCAGGGAGCGCTTCTCTCCTCGATAGCCCTGTGGCGAGCTTCATGCCT 1224
      |||||||

QY 380 CCTACTGGCGCTACGCCAACTCCTG----- 405
      |||||||
Db 1225 CCTACTGGCGCTACGCCAACTCCTCGAGGTATGAAGTAGGGCGTGGAGATGGGGCT 1284
      |||||||

QY 406 ----- 405
      |||||||
Db 1285 TGCAGGTGTCAGAGACAGAGGGTTGGGGTTAGGGTTAGAGTCTTCTGACTGTGC 1344
      |||||||

QY 406 -----CAGCCTGAGGGTCACCACTGGGAGACTCAGCAGATTCGAAGCCTCAGTC 454
      |||||||
Db 1345 CTATGTCCTTCAGCCTGAGGGTCACCACTGGGAGACTCAGCAGATTCGAAGCCTCAGTC 1404
      |||||||

QY 455 CCAGCCAGCCATGGCAGCGTCTCTTCTCGGTTCAAAATCCTTCGACGCCCTCCAGGCCT 514
      |||||||
Db 1405 CCAGCCAGCCATGGCAGCGTCTCTTCTCGGTTCAAAATCCTTCGACGCCCTCCAGGCCT 1464
      |||||||

QY 515 TTGTGGCTGTAGCCGCCCGGCTTTGCCCATGGAGCAGCAACCTGATCCCTAA 570
      |||||||
Db 1465 TTGTGGCTGTAGCCGCCCGGCTTTGCCCATGGAGCAGCAACCTGATCCCTAA 1520
      |||||||

RESULT      8
AC025574/c
LOCUS       AC025574 155929 bp DNA HTG 06-SEP-2000
DEFINITION Homo sapiens chromosome 12 clone RP11-348M3, WORKING DRAFT
SEQUENCE, 14 unchromosome pieces.
ACCESSION  AC025574
VERSION    AC025574.10 GI:9438393
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 155929)
            Muzny D.M., Adams C., Bailey M., Barberia J., Blankenburg K.,
            Bodota B., Bouck J., Bowie S., Brooks A., Buhay C., Bunac C.,
            Burkett C., Burrows J., Carter M., Chacko J., Chen Z., Cox C.,
            David R., Delgado O., Deshazo D., Ding Y., Domah-Rashid N.,
            Dugan-Rocha S., Durbin K.J., Fernandez C., Ferraguto D.,
            Forcum-Tansey J., Frantz P., Ganesh R., Correll J.H., Correll L.L.,
            Guevara W., Harris K., Hernandez J., Hodgson A., Hogue M.,
            Holloway C., Hosak H., Jackson L.E., Jackson L., Jia Y., Jones M.,
            Kelly S., Kondejewski N., Kong Y., Kovar C., Leal B., Li Z.,
            Lichtarge O., Liu J., Liu W., Logan O., Lozado R.J., Lu J.,
            Lucier R., Martin R., Martinez C., McLeod M.P., Mei G., Morgan M.,
            Morris S., Nash S., Nelson A., Nguyen R., Nguyen N., Nguyen S.,
            Oswal G., Parish B., Paxton S., Payton B., Perez L., Pu L.L.,
            Quiles M., Reiter D., Rives M., Samuel S., Say J., Scherer S.,
            Shah E., Shen H., Simon M., Sparks A., Stamps A., Sucgang R.,
            Tabor P., Taylor T., Vasquez L., Vinson R., Vo O., Wabbah M.,
            Watlington S., Weinstock G., Weinstock I.R., Williamson A., and
            Worley K., Wren J., Wrensford G., Yu W., Zhou X., Nelson D., and
            Gibbs R.
Direct Submission
Unpublished
2 (bases 1 to 155929)
Worley K.C.
Submitted (11-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 25, 2000 this sequence version replaced gi:8571540.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HALV
Center clone name: RP11-348M3
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodypy: 6% of reads
Chemistry: Dye-terminator Big Dye: 94% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 133622 bases at least Q40
Consensus quality: 145733 bases at least Q30
Consensus quality: 150334 bases at least Q20
Estimated insert size: 150694; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 36972: contig of 36972 bp in length
* 36973 37072: gap of unknown length
* 37073 64901: contig of 27829 bp in length
* 64902 65001: gap of unknown length
* 65002 91014: contig of 26013 bp in length
* 91015 91114: gap of unknown length
* 91115 105122: contig of 14008 bp in length
* 105123 105222: gap of unknown length
* 105223 113963: contig of 8741 bp in length
* 113964 114064: gap of unknown length
* 114064 124894: contig of 10831 bp in length
* 124895 124994: gap of unknown length
* 124995 132612: contig of 7618 bp in length
* 132613 132712: gap of unknown length
* 132713 139659: contig of 6947 bp in length
* 139660 139759: gap of unknown length
* 139760 143764: contig of 4005 bp in length

```

\* 143765 143864: gap of unknown length  
 \* 143865 148062: contig of 4198 bp in length  
 \* 148063 148162: gap of unknown length  
 \* 148163 151000: contig of 2838 bp in length  
 \* 151001 151100: gap of unknown length  
 \* 151101 153374: contig of 2274 bp in length  
 \* 153375 153474: gap of unknown length  
 \* 153475 154810: contig of 1336 bp in length  
 \* 154811 154910: gap of unknown length  
 \* 154911 155929: contig of 1019 bp in length.

FEATURES  
 source  
 1. .155929  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="12"  
 /clone="RP11-348M3"

BASE COUNT 40876 a 36387 c 36205 g 41119 t 1342 others  
 ORIGIN

Query Match 33.8%; Score 192.8; DB 69; Length 155929;  
 Best Local Similarity 74.3%; Pred. No. 1.6e-44;  
 Matches 309; Conservative 0; Mismatches 2; Indels 105; Gaps 1;

QY 260 AGTTCGCTGCAAGAGTCCACAGGGTCTGATTTTATGAGAAGCTGCTAGGATCGG 319  
 Db 63431 AGTTCGCTGCAAGAGTCCACAGGGTCTGATTTTATGAGAAGCTGCTAGGATCGG 63372  
 QY 320 ATATTTTACAGGGAGCCCTTCTCTCCCTGATAGCCCTGTGGCGCATTCATGCCCT 379  
 Db 63371 ATATTTTACAGGGAGCCCTTCTCTCCCTGATAGCCCTGTGGCGCATTCATGCCCT 63312  
 QY 380 CCTACTGGCCCTCAGCAACTCTCTG----- 405  
 Db 63311 CCTACTGGCCCTCAGCAACTCTCTGAGTATGAAGTAGGGCGGTGGAGATGGGGGT 63252  
 QY 406 ----- 405  
 Db 63251 TGCAGGTGTGACAGACAGAGGGTTGGGGTTAAGGGTTTAGAGTCTCTGACTGTGTC 63192  
 QY 406 -----CAGCCTGAGGGTACACTGGGAGACTCAGCAGATTCGAAGCCTCAGTC 454  
 Db 63191 CTATGCTCTTTCAGCCTGAGGGTTCAGCACTGGGAGACTCAGCAGATTCGAAGCCTCAGTC 63132  
 QY 455 CCAGCCAGCATGCGAGCGTCTCTCTCGCTTCAAAATCTTCGAGCCTCCAGCCCT 514  
 Db 63131 CCAGCCAGCATGCGAGCGTCTCTCTCGCTTCAAAATCTTCGAGCCTCCAGCCCT 63072  
 QY 515 TTGTGGCTGTAGCGCCCGGGTCTTTGCCCATGGAGCAGCAACCTGAGTCCCTAA 570  
 Db 63071 TTGTGGCTGTAGCGCCCGGGTCTTTGCCCATGGAGCAGCAACCTGAGTCCCTAA 63016

RESULT 9  
 AB030002  
 LOCUS AB030002 2300 bp DNA MAM 04-APR-2000  
 DEFINITION Sus scrofa gene for SGRF, complete cds.  
 ACCESSION AB030002  
 VERSION AB030002.1 GI:7416076  
 KEYWORDS SGRF.  
 SOURCE Sus scrofa  
 ORGANISM Sus scrofa

REFERENCE  
 AUTHORS Hirata,Y. and Kosuge,Y.  
 TITLE SGRF: a novel member of the IL-6/g-CSF family  
 JOURNAL Published only in DataBase (2000) In press  
 REFERENCE  
 AUTHORS Hirata,Y. and Kosuge,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-JUL-1999) to the DDBJ/EMBL/GenBank databases. Yuichi Hirata, Chugai Research Institute for Molecular Medicine, Gene

Search Program: 153-2 Nagai, Niihari-mura, Ibaraki 300-4101, Japan  
 (E-mail:hiratayus@immed.com, Tel:81-298-30-6211(ex.288),  
 Fax:81-298-30-6270)

FEATURES  
 source  
 1. .2300  
 Location/Qualifiers

/organism="Sus scrofa"  
 /db\_xref="taxon:9823"

TATA\_signal  
 227..232

gene  
 Join(433..603,820..921,1590..1736,1837..1998)

CDS  
 Join(433..603,820..921,1590..1736,1837..1998)

/gene="SGRF"  
 /codon\_start=1  
 /product="SGRF"

/protein\_id="BAA93688.1"  
 /db\_xref="GI:7416077"

/translation="MLGSRAVLMMLLLPMTSGRAVPESSPAWAGQOOLSOOLC  
 TLAWTAHLPMGHVDLPREGGDDTTSEVPHIQCGDGDQGLRDSQSLQRIHGLV  
 FYEKLLGSDIFTEPSLHPDGSVGLHSLGLRLQLLOPEGHMETEQTPSPSPQPW  
 ORLLRLKILRSLOAFVAVARVFAHGAATLSQ"

polyA\_signal 2265..2270  
 BASE COUNT 597 a 563 c 589 g 551 t

ORIGIN

Query Match 26.2%; Score 149.2; DB 7; Length 2300;  
 Best Local Similarity 67.7%; Pred. No. 6e-32;  
 Matches 279; Conservative 0; Mismatches 33; Indels 100; Gaps 1;

QY 259 CAGTCTGCTTGCACAGGATCCACAGGGTCTGATTTTATGAGAAGCTGCTAGGATCG 318  
 Db 1587 CAGTCTGCTTGCACAGGATCCACAGGGTCTGATTTTATGAGAAGCTGCTAGGATCG 1646  
 QY 319 GATATTTTACAGGGAGCCCTTCTCTCCCTGATAGCCCTGTGGCGCAGCTTCATGCC 378  
 Db 1647 GACATTTTACAGGGAGCCCTTCTCTACACCCCTGATGGCTCTGTGGCCAGCTTCAGCC 1706  
 QY 379 TCCTACTGGCCCTCAGCCCAACTCTG----- 405  
 Db 1707 TCCTACTGGCCCTCAGCCCAACTCTTGCAGGTATAAACTAGGGCCCTGGAGCGGGGGTG 1766  
 QY 406 ----- 405  
 Db 1767 GCTTCAGGTGTGACAGACAGTGCCTGGGGGTGAAGATCCTCTCTACTAGTGTCTGT 1826  
 QY 406 -----CAGCCTGAGGGTACACCTGGGAGACTCAGCAGATTCGAAGCCTCAGTCCCAG 458  
 Db 1827 GTCTTTTACAGCGAGGGTCAACCTGGGAGACTGAGCAGACGCCAAGCCCAAGTCCCAG 1886  
 QY 459 CCAGCCATGCGAGCGTCTCTCTCCCTTCAAAATCTTCGAGCCTCCAGCCCTTGT 518  
 Db 1887 CCAGCCATGCGAGCGTCTCTCTCCCTTCAAAATCTTCGAGCCTCCAGCCCTTGT 1946  
 QY 519 GGCTGTAGCGCCCGGGTCTTTGCCCATGGAGCAGCAACCTCAGTCCCTAA 570  
 Db 1947 GGCTGTAGCGCCCGGGTCTTTGCCCATGGAGCAGCAACCTCAGTCCCTAA 1998

RESULT 10  
 AB004061/c  
 LOCUS AB004061 4649 bp mRNA MAM 05-FEB-1999  
 DEFINITION domestic pig mRNA for STAT2, complete cds.  
 ACCESSION AB004061  
 VERSION AB004061.1 GI:2189979  
 KEYWORDS STAT2.  
 SOURCE Sus scrofa adult muscle thoracic aortic cell cDNA to mRNA.  
 ORGANISM Sus scrofa

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Cetartiodactyla; Suidae; Sus.  
 TITLE Ito,Y.  
 JOURNAL Direct Submission  
 Submitted (20-MAY-1997) to the DDBJ/EMBL/GenBank databases.

Yoshiyasu Ito, National Institute of Animal Industry, Animal Genome Research Team; 2 Ikenodai, Kuki-zaki-machi, Inashiki-gun, Ibaragi 305-0901, Japan (E-mail:yoshiito@niai.affrc.go.jp, Tel:81-298-38-8627)

2 (sites)  
 Ito, Y., MIKAWA, S., Kobayashi, E., Wada, Y. and MINEZAWA, M.  
 PIG STAT2 cDNA  
 Unpublished (1997)  
 Location/Qualifiers  
 1..4649  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /cell\_type="thoracic aortic cell"  
 /dev\_stage="adult"  
 /tissue\_type="muscle"  
 82..2676  
 /codon\_start=1  
 /product="STAT2"  
 /protein\_id="BAA20332.1"  
 /db\_xref="GI:2189980"  
 /translation="MAQWEMQLNDSPFDQLHOLYSESLPVDVROYLAVWLEDQNW  
 OEALGNDSDKANLPEFHFDOLNYDCGCGQDEPCILLLOHNLKRYRDIAIPQOPT  
 RLAMFNLLEKRLILQARQAEQEPALAPGENQOHEIESRILELRAMEKLY  
 KSLQKDDQDIFCFRYKIRASAKTHSLDPRHQEQVLOETNELDKRKEVLDASK  
 ALIGRLTLELLPKLEWKVQOQKACIGAPMDGELEQEKWFTAPAKLLFHLROL  
 KELGSSVYKRLDDELLFKGVDLLKQVTELLHRAPIVETQPCMPQTPRPLIL  
 KTGKFTVRLRLVROEGNESLTAESIDRNPPKSGQFRKNILTSNRKTLTPKQGO  
 SOGLINDFQVLTMEORSAGGKNNKPLGVTEELHIIISFTVKYNYGQKQLTDT  
 LPVITISNMQLSIWASILWFLNLSDDPQOQFFSPKAPWNLLGPAISWQFSSHY  
 GQGLNDQMLKRLFGQNSSTEGLSWVDFIKRESPPKLPFWTWLKDILDLVHD  
 HKLDKMDHMGFVSSEERLLKTTISGTLFRSETPLEGGITCSWHEQDDKVL  
 IYSLQPTKEVLQSLPLTKIISQYQLTEENIPENFLYPRIPRDEAFEGYDQKKA  
 NLQERKYLKLEPLTVSNQVDELOQPELKLDPDLSELDLGLAPGPPGVLGLE  
 PLEAGLDLQPELMLQSTLEPVLDELQVPEPPLGPELKLPLEPTELVSPQV  
 PEPDLPHLHNTDENQIFRNSMRIEIMPNGDPLLASQNTVNDENCFHRSHFYTD  
 GPLIPSDY"

polyA\_site 3622  
 /note="Alternative Poly adenilation site"  
 polyA\_site 4649  
 /note="21 a nucleotides"  
 BASE COUNT 1234 a 1157 c 1194 g 1064 t  
 ORIGIN  
 Query Match 26.2%; Score 149.2; DB 7; Length 4649;  
 Best Local Similarity 67.7%; Pred. No. 5.8e-32;  
 Matches 279; Conservative 0; Mismatches 33; Indels 100; Gaps 1;  
 Qy 259 CAGTCTGCTGCAAGAGATCCACAGGCTGTGATTTTATGAGAGCTGCTAGGATCG 318  
 Db 4109 CAGTCTGCTGCAAGAGATCCACAGGCTGTGATTTTATGAGAGCTGCTAGGATCG 4050  
 Qy 319 GATATTTTCAGGGGAGCCCTTCTGCTCCCTGATAGCCCTGTGGCGACGCTCATGCC 378  
 Db 4049 GATATTTTCAGGGGAGCCCTTCTGCTCCCTGATAGCCCTGTGGCGACGCTCATGCC 3990  
 Qy 379 TCCCTACTGGCCCTCAGCAACTCCG----- 405  
 Db 3989 TCCCTACTGGCCCTCAGCAACTCTTGCAAGTATATAAATAGGGGCTTGACGGGGGTG 3930  
 Qy 406 ----- 405  
 Db 3929 GCTTGCAGGTGTGAGAGACATGCTGGGGGTGAAGGATCCTCTACTAGTGTCTGTG 3870  
 Qy 406 -----CAGGCTGAGGTCACCACTGGGAGACTCAGCAGATTCACAGCCCTCAGTCCAG 458  
 Db 3869 GTCCTTTTCAGGCCGAGGCTCACCCTGAGGAGCTGAGCAGGACGACCCAGTCCAG 3810  
 Qy 459 CCAGCCATGGCAGGCTCTCTCTCCGCTTCAAAATCTTCGCGACGCTCCAGGCTTTGT 518  
 Db 3809 CCAGCCCTGGCAGCGCTCTCTCTCCGCTTCAAGATCTTCCGCGACGCTCCAGGCTTTGT 3750  
 Qy 519 GCCTGTAGCCGCCCGGGTCTTTGGCCCTGAGGAGCAACCTGTAGTCCCTAA 570

||||| GCGTGTAGTCCCGGGTCTTCGCCCGGAGCAGCAACTCTGAGCCAGTAA 3698  
 Db 3749  
 RESULT 11  
 SCPSCL 11696 bp DNA BCT 30-JUN-1993  
 LOCUS S.clavuligerus linear plasmid pSCL (complete sequence).  
 DEFINITION X54107.1 GI:48758  
 ACCESSION X54107.1  
 VERSION X54107.1  
 KEYWORDS inverted repeat; linear plasmid.  
 SOURCE Streptomyces clavuligerus.  
 ORGANISM Streptomyces clavuligerus  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 REFERENCE 1 (bases 1 to 11696)  
 AUTHORS Roy, K.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-JUL-1990) Roy K.L., Dept of Microbiology, University of Alberta, Biological Sciences Bldg Rm M330, Edmonton Alberta, Canada T6G 2E9  
 REFERENCE 2 (bases 1 to 11696)  
 AUTHORS Wu, X. and Roy, K.L.  
 TITLE Complete nucleotide sequence of a linear plasmid from Streptomyces clavuligerus and characterization of its RNA transcripts  
 JOURNAL J. Bacteriol. 175 (1), 37-52 (1993)  
 MEDLINE 93106972  
 FEATURES Location/Qualifiers  
 source 1..11696  
 /organism="Streptomyces clavuligerus"  
 /plasmid="pSCL"  
 /strain="NRRL 3585"  
 /db\_xref="taxon:1901"  
 complement(1446..2237)  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA38040.1"  
 /db\_xref="GI:48759"  
 /translation="MAMTRDALLNQGGVAKTTTLLHGGTLLAAGRTLLVLDGOG  
 NLTTALKPLRNAGELTLARVLEGCDRAGASLVRHRSNENLWYIPSAOLDTLPQ  
 LHSARKEERLSWLEHLEDGFDHCLVDVDPDNDVLRWADHLLIPVDVDEFSI  
 EALELLIGQVTLVRETRMAPPYRGVLRVIRVAPFSGFQAVYDALHRLPLPVGSI  
 PLRTAVAEAKNQGTQAQYARTADVARMERDLAVKAGYLPEAPAA"  
 complement(4062..5417)  
 /note="ORF2"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA38041.1"  
 /db\_xref="GI:581632"  
 /db\_xref="SPTREMBL:Q05073"  
 /translation="MAMTRDALLNQGGVAKTTTLLHGGTLLAAGRTLLVLDGOG  
 NLTTALKPLRNAGELTLARVLEGCDRAGASLVRHRSNENLWYIPSAOLDTLPQ  
 LHSARKEERLSWLEHLEDGFDHCLVDVDPDNDVLRWADHLLIPVDVDEFSI  
 EALELLIGQVTLVRETRMAPPYRGVLRVIRVAPFSGFQAVYDALHRLPLPVGSI  
 PLRTAVAEAKNQGTQAQYARTADVARMERDLAVKAGYLPEAPAA"  
 complement(4062..5417)  
 /note="ORF2"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA38041.1"  
 /db\_xref="GI:581632"  
 /db\_xref="SPTREMBL:Q05074"  
 /translation="MSMESEELRAAEVREAEADAVLGTLLRPEARLDVPLVLSALPTD  
 FRPRYGEMLDLSLWARTPCDPVTYSVLRKGDRLRIGGNVHLHLDVPTTA  
 HALRYAEIVAEESRRALLAAGVTVTAQRREGTAETAEATAGVAVREVDRGLAAH  
 DAPYALDIDLDQADDEPEWIFGVLAANDRLMTAGEGGKSLQLLRLVRAAGLH  
 PWRARIEPVYVLLVDAENASVORPWLRRMAAAGAEVARGRMHIEVRPEGLD  
 TTAADRAWLARTVERVPDLMSTIGPLYKLTTSGSDSEAAARPLMSALEMVRQASDGA  
 MLEAHSPPHAPGVRRLDRPLRIGSSLWRPFGGLSPATGMGAELRLMDLWLPWRG  
 PRSERWPAQFCGGTWPVAVQYAGNVPIPAGTLTTEQAVAGLIPQEPFFWDDQOI  
 G"  
 complement(6349..6786)  
 /note="ORF5"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA38042.1"  
 /db\_xref="GI:48761"  
 /db\_xref="SPTREMBL:Q05076"  
 /translation="MEAAHLRFAQLRQHGEWDFGSDNPVLKVEEAITAGLPEIPAQ  
 QPSHRVIPPSPSVSKDARCGHLSFHRGTAGTAVAGWDLDCSCERFRAEADWH  
 VGMPCFCHRSADHLGIEGCSGKGCACCCYHFSASGVGAR"  
 complement(6931..7914)  
 /note="large RNA transcript"  
 prim\_transcript

```

CDS
complement(7139..7885)
/codon_start=1
/transl_table=11
/protein_id="CAA38043.1"
/db_xref="GI:48762"
/db_xref="SPTREMBL:Q05077"
/translation="MPKIERPLPPYMQVIAHRDQIDSGELRAGDLIPSDRNIAQEWG
ISRTAQKVTALKAQGLVEVQSGTRVRAAAADHLHRSGRDRRAASTVRRGTGRITYTAGE
YARISAEIAPADPAADVAELSVQEGAAVIRIRITVYNAENOPISASTSWDGNLAAGA
PKLLKTERIKESGWAYLESGQTGTAIVYGODRISARMAEQDAQLLGLIELPAALKEART
ILRDANSVAVEGVSINGNGRSIYDVS"
promoter
complement(7921..7951)
/note="promoter for large RNA transcript"
prim_transcript
complement(7951..8400)
/note="small RNA transcript"
CDS
8114..9367
/note="ORF3"
/codon_start=1
/transl_table=11
/protein_id="CAA38044.1"
/db_xref="GI:581633"
/db_xref="SPTREMBL:Q05075"
/translation="MLEAPRPUSDPTCLARRSDVSIVARRVVAQGVPRVPVAGLAEQ
VEWEMREDDAQGTDPQEGGVKMSQRRELGAARRAVVPGQRRVTPSTAPDRVDLR
TABDVELLEVPDLRVRAALVAEEHGDHLVYVAARLDKAWAYWARAEDSAQDVWLDVLR
GRVPELLEVPDLRWPLRLAAAKWSLDNTRRRRLRVPRPQREVLELRCTEGMTTPAIAR
GPDSTVCANDELMPEPEPGGWAPACYABRIALPPQREVLELRCTEGMTTPAIAR
LGLRSOSVDRALRFATLALGSPGVVVRSGRQGPAGWERNVLDLRLPNRQORDVVR
LRAGASFGELGGLHGRGYAHLEYTRALRSUREMVQDRLDPVPAAPRSKPKP
APAGCAATSCVSGCYLRTARTGAAS"
promoter
complement(8411..8442)
/note="promoter for small RNA transcript"
CDS
9696..10679
/note="ORF4"
/codon_start=1
/transl_table=11
/protein_id="CAA38045.1"
/db_xref="GI:48764"
/db_xref="SPTREMBL:Q05083"
/translation="MGREETPGVPRQWVLRLLGGAVAVGAVAMAASAVTLAELGHA
VGVQWMDRLSWSLPAVDLLALVAGVWLSPGMTARSLRGITLGAVGSVALNIA
IGHLVESGDMKVTPLLRIAVSAVPLVAALVHLVGLVLPAGDHGQDQAVLELEQ
REAAITSAADHPADGTTADRHGCTTARNAPAPQRPAPADAPVRVPAPAAQADARTAA
DDTPAPGVPAAAGTGPRTVRLGDRPDDSOIWERPAPVRQWDRPDLADCAPGTEP
LLLTPODSRARILWALQNTDWTQKRIGEFAGRSASVYVNRKAEAIRDGILI"
BASE COUNT 1590 a 4235 c 4179 g 1692 t
ORIGIN

Query Match 7.2% Score 40.8; DB 3; Length 11696;
Best Local Similarity 51.7%; Pred. No. 0.66; Indels 0; Gaps 0;
Matches 93; Conservative 0; Mismatches 87;

Qy 336 GCCTTCTGTCCTGATAGCCCTGTGTGGGCAGCTTCATCGCTCCCTATGGGCCCTCAG 395
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5222 GCGTCCCGCGGCCACAGCGACCCAGCAGATCCACAACTCCCGCTACCGGGCTCCC 5281

Qy 396 CCAACTCCTGACGCTGAGGCTACCACTGGGAGACTCAGACAGATCCAGCCTCAGTCC 455
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5282 GGAAATCCGTGGGGCGGAGCGGAGACCAAGATCCGCGACAGCAGATCCCGCGCTCCGCGC 5341

Qy 456 CAGCAGCATGGCAGGCTCTCTTCCTCCGCTTCAAAATCCTTCGACGCTCCAGGCTT 515
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5342 GCAGCAGCTGCCAGCAGCCGCTCTCGGCTCTCCCGACCTCAGCCGCGCAGCTCCT 5401

RESULT 12
CNS06SJ1/c 937 bp DNA STS 10-JAN-2001
LOCUS T3 end of clone AW0AA023f10 of library AW0AA from strain CLIB 89 of
DEFINITION Yarrowia lipolytica, sequence tagged site.
ACCESSION AL413315

```

**RESULT** 13  
**HS355C18**  
**LOCUS**  
**DEFINITION**  
 Human DNA sequence from clone RP3-355C18 on chromosome 22q13.3  
 Contains the KIAA0027 gene, ESTs, GSSs and seven putative CpG  
 islands, complete sequence.  
**ACCESSION**  
 AL022327  
**VERSION**  
 AL022327.17 GI:5304851  
**KEYWORDS**  
 HTG; CpG island; KIAA0027.  
**SOURCE**  
 human.  
**ORGANISM**  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE**  
**AUTHORS**  
 Cobley, V.  
**TITLE**  
 Direct Submission  
**JOURNAL**  
 Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: [hkquery@sanger.ac.uk](mailto:hkquery@sanger.ac.uk)  
 requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Jun 30, 1999 this sequence version replaced gi:5262834.  
**COMMENT**  
 This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.  
 This sequence was generated from part of bacterial clone contigs of  
 human chromosome 22, constructed by the Sanger Centre Chromosome 22  
 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr22>  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)  
 from the library RPCI-3 constructed at the Roswell Park Cancer  
 Institute by the group of Pieter de Jong. For further details see  
<http://bacpac.med.buffalo.edu/>  
 VECTOR: pCYPAC2  
 This sequence is the entire insert of clone RP3-355C18 The true  
 left end of clone RP5-89814 is at 5390 in this sequence.  
**FEATURES**  
 source  
 Location/Qualifiers  
 1..101270  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /map="q13.3"  
 /clone="RP3-355C18"  
 /clone.lib="RPCI-3"  
 1..1038  
 /note="match: STS: Em:AL022483; match: STS: Em:HS355C18T"  
 repeat\_region  
 1..242  
 /note="AluJo repeat: matches 1..237 of consensus"  
 repeat\_region  
 247..322  
 /note="L1 repeat: matches 3067..3140 of consensus"  
 repeat\_region  
 358..484  
 /note="L1M4c repeat: matches 816..940 of consensus"  
 repeat\_region  
 485..792  
 /note="AluX repeat: matches 2..305 of consensus"  
 repeat\_region  
 795..1095  
 /note="AluX repeat: matches 1..299 of consensus"  
 repeat\_region  
 1096..1163  
 /note="L1M4c repeat: matches 940..1017 of consensus"  
 repeat\_region  
 1165..1551  
 /note="MSTB repeat: matches 3..412 of consensus"  
 repeat\_region  
 1564..1750  
 /note="L1M4c repeat: matches 1012..1190 of consensus"  
 repeat\_region  
 1757..2038  
 /note="AluX repeat: matches 1..282 of consensus"  
 repeat\_region  
 2054..3342  
 /note="SVA repeat: matches 1..1386 of consensus"  
 repeat\_region  
 3343..3412  
 /note="35 copies 2 mer tg 82 conserved"  
 repeat\_region  
 3418..3577  
 /note="L1M4c repeat: matches 1190..1350 of consensus"  
 repeat\_region  
 3578..3861  
 /note="AluJo repeat: matches 1..300 of consensus"  
 repeat\_region  
 3862..4334  
 /note="L1M4c repeat: matches 1350..1817 of consensus"  
 repeat\_region  
 4341..4489  
 /note="L1M4c repeat: matches 2052..1812 of consensus"  
 repeat\_region  
 4492..4858  
 /note="MER4A repeat: matches 237..650 of consensus"  
 repeat\_region  
 5142..5300  
 /note="AluJb repeat: matches 137..295 of consensus"  
 repeat\_region  
 5315..5623  
 /note="AluX repeat: matches 3..312 of consensus"  
 repeat\_region  
 5626..5786  
 /note="L2 repeat: matches 2492..2665 of consensus"  
 repeat\_region  
 11158..11247  
 /note="45 copies 2 mer tg 68 conserved"  
 repeat\_region  
 11251..11320  
 /note="35 copies 2 mer tg 72 conserved"  
 repeat\_region  
 11369..11690  
 /note="161 copies 2 mer tg 73 conserved"  
 repeat\_region  
 11372..11679  
 /note="7 copies 44 mer 72 conserved"  
 complement(11521)  
 /note="Tandem repeat. Assembly not confirmed by digest"  
 11973..13056  
 /note="CpG island"  
 /evidence=not\_experimental  
 12226..12498  
 /note="13 copies 21 mer 63 conserved"  
 repeat\_region  
 14952..15097  
 /note="MER51B repeat: matches 443..601 of consensus"  
 repeat\_region  
 15195..15503  
 /note="AluX repeat: matches 1..311 of consensus"  
 repeat\_region  
 15540..15948  
 /note="L1M4c repeat: matches 292..736 of consensus"  
 repeat\_region  
 15952..16252  
 /note="AluX repeat: matches 1..306 of consensus"  
 repeat\_region  
 16253..16297  
 /note="Alu repeat: matches 1..48 of consensus"  
 repeat\_region  
 16322..16780  
 /note="L1M4c repeat: matches 3883..4321 of consensus"  
 repeat\_region  
 16781..17050  
 /note="AluX repeat: matches 38..307 of consensus"  
 repeat\_region  
 17051..17106  
 /note="L1M4c repeat: matches 4321..4378 of consensus"  
 repeat\_region  
 17309..17576  
 /note="AluX repeat: matches 22..298 of consensus"  
 repeat\_region  
 17856..17701  
 /note="23 copies 2 mer aa 80 conserved"  
 repeat\_region  
 18229..18553  
 /note="AluX repeat: matches 1..312 of consensus"  
 repeat\_region  
 19381..19680  
 /note="AluJo repeat: matches 1..299 of consensus"  
 repeat\_region  
 19700..19923  
 /note="L1 repeat: matches 3883..4106 of consensus"  
 repeat\_region  
 20070..20379  
 /note="AluX repeat: matches 9..310 of consensus"  
 repeat\_region  
 20416..20470  
 /note="U6 repeat: matches 48..102 of consensus"  
 repeat\_region  
 20666..20821  
 /note="Alusq/x repeat: matches 1..137 of consensus"  
 repeat\_region  
 21593..21916  
 /note="L2 repeat: matches 2435..2724 of consensus"  
 repeat\_region  
 22848..23801  
 /note="CpG island"  
 misc\_feature

```
repeat_region /evidence-not_experimental
23480. .23537
/notes="29 copies 2 mer 99 72 conserved"
repeat_region 23650. .23804
/notes="5 copies 31 mer 66 conserved"
repeat_region 23654. .24010
/notes="17 copies 21 mer 58 conserved"
repeat_region 26256. .26317
/notes="31 copies 2 mer 98 conserved"
misc_feature 27337. .28417
/notes="CpG island"
/evidence-not_experimental
28029. .28422
/notes="197 copies 2 mer cc 56 conserved"
repeat_region 28038. .28436
/notes="19 copies 21 mer 56 conserved"
repeat_region 28606. .28661
/notes="28 copies 2 mer cc 71 conserved"
repeat_region 30789. .30970
/notes="FRAM repeat: matches -8. .174 of consensus"
repeat_region 30989. .31124
/notes="L1M4c repeat: matches 1258. .1422 of consensus"
repeat_region 31187. .31308
/notes="L1M4c repeat: matches 1525. .1649 of consensus"
misc_feature join(31289. .31500,43139. .43315)
/notes="match: STS: Em:A1026623"
repeat_region 31510. .32051
/notes="L1M4 repeat: matches 1881. .2118 of consensus"
repeat_region 32054. .32334
/notes="MLTIF repeat: matches 1. .536 of consensus"
repeat_region 32541. .32908
/notes="L1M4 repeat: matches 2129. .2499 of consensus"
repeat_region 33133. .33264
/notes="3 copies 44 mer 74 conserved"
repeat_region 33279. .33410
/notes="3 copies 44 mer 85 conserved"
misc_feature complement(33332)
/notes="Random repeat Forced join Each element 22 bases,
typical sequence: GACACGTGCACCCCTCATCT. Restriction
digest data (ECORI, HINDIII) suggest around 4 copies of
the repeat are missing from this assembly."
repeat_region 34730. .34817
/notes="2 copies 44 mer 86 conserved"
repeat_region 34844. .34931
/notes="2 copies 44 mer 95 conserved"
repeat_region 35030. .35117
/notes="2 copies 44 mer 95 conserved"
repeat_region 35455. .35648
/notes="L1M4 repeat: matches 4149. .4351 of consensus"
repeat_region 35733. .36111
/notes="L1M4 repeat: matches 4263. .4643 of consensus"
repeat_region 36112. .36420
/notes="AluSc repeat: matches 1. .307 of consensus"
repeat_region 36421. .36447
/notes="L1M4 repeat: matches 4643. .4669 of consensus"
repeat_region 36448. .36756
Query Match 7.0%; Score 40; DB 92; Length 101270;
Best Local Similarity 61.5%; Pred.No.1;
Matches 64; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 36 GCTGCCTGCAGCTCAGGGCAGAGTGTGCTGTGGGCGACAGCCCTCGCTGGACTCA 95
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61606 GTTGGCCAGGCCCTCTGTCACACACACACAGCCGGGCTGTCCGAGCCCTGCCGGCCCTCA 61665
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 96 GTGCCAGCAGCTTTCACAGAAGCTGTGCACACTGCGCTGGAGTG 139
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61666 GTGCCTCAGCTTCCACATGCCACAGCGCCAGCCAGTGTG 61709
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 14
AC022220
LOCUS AC022220 233231 bp DNA HTG 17-AUG-2000
```

```
DEFINITION Homo sapiens chromosome X clone RP11-815F17, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
ACCESSION AC022220
VERSION AC022220.7 GI:9838350
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 233231)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 233231)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 17, 2000 this sequence version replaced gi:8980041.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0815F17
----- Summary Statistics -----
Sequencing vector: M13; 66%
Chemistry: Dye-primer ET; 66% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 228369 bases at least Q40
Consensus quality: 229897 bases at least Q30
Consensus quality: 230504 bases at least Q20
Insert size: 214000; agarose-fp
Insert size: 233559; sum-of-contigs
Quality coverage: 7.60 in Q20 bases; agarose-fp
Quality coverage: 6.98 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1224: contig of 1224 bp in length
* 1225 1324: gap of unknown length
* 1325 11218: contig of 9894 bp in length
* 11219 11318: gap of unknown length
* 11319 20516: contig of 9198 bp in length
* 20517 20616: gap of unknown length
* 20617 34259: contig of 13643 bp in length
* 34260 34359: gap of unknown length
* 34360 47343: contig of 12984 bp in length
* 47344 47443: gap of unknown length
* 47444 63537: contig of 16094 bp in length
* 63538 63637: gap of unknown length
* 63638 91959: contig of 28322 bp in length
* 91960 92059: gap of unknown length
* 92060 127451: contig of 33391 bp in length
* 127451 182595: gap of unknown length
* 182595 182696: gap of unknown length
* 182696 233231: contig of 50536 bp in length.
* Location/Qualifiers
* 1..233231
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="X"
```











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 00:52:31 ; Search time 153.2 seconds  
(without alignments)  
2172.032 Million cell updates/sec

Title: US-09-558-474-1  
Perfect score: 570  
Sequence: 1 ATGCTGGGAGCAGAGCTGT.....CAGCAACCTGAGTCCTTAA 570

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues  
Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0401.\*  
1: /SID32/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SID32/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SID32/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SID32/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SID32/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SID32/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SID32/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SID32/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SID32/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SID32/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SID32/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SID32/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SID32/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SID32/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SID32/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SID32/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SID32/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SID32/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SID32/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SID32/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SID32/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SID32/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Query Match | Score | Length | ID | Description               |
|------------|-------------|-------|--------|----|---------------------------|
| 1          | 570         | 100.0 | 570    | 20 | 208865 Human interleukin  |
| 2          | 570         | 100.0 | 570    | 20 | xi17786 Human interleukin |
| 3          | 570         | 100.0 | 570    | 21 | A52577 Human interleukin  |
| 4          | 566.8       | 99.4  | 1026   | 21 | 237262 SGFR coding sequen |
| 5          | 566.8       | 99.4  | 1055   | 22 | C84306 Human EXCS encodin |
| 6          | 566.8       | 99.4  | 1067   | 21 | A16686 Human secreted pro |
| 7          | 391         | 68.6  | 1203   | 20 | 208866 Mouse interleukin  |
| 8          | 391         | 68.6  | 1203   | 20 | xi17787 Mouse interleukin |
| 9          | 391         | 68.6  | 1203   | 21 | A52578 Murine interleukin |
| 10         | 242.8       | 42.6  | 412    | 21 | A42676 Human secreted exp |
| 11         | 192.8       | 33.8  | 2398   | 21 | 237263 SGFR coding sequen |

|   |    |      |     |        |    |        |                    |
|---|----|------|-----|--------|----|--------|--------------------|
| c | 12 | 38   | 6.7 | 121162 | 21 | C66548 | Human kinesin-like |
| c | 13 | 35   | 6.1 | 567    | 21 | Z55567 | Feline interferon- |
| c | 14 | 35   | 6.1 | 567    | 21 | Z55568 | Feline interferon- |
| c | 15 | 35   | 6.1 | 567    | 21 | Z55569 | Feline interferon- |
| c | 16 | 35   | 6.1 | 567    | 21 | Z55570 | Feline interferon- |
| c | 17 | 35   | 6.1 | 582    | 10 | N90186 | Feline interferon- |
| c | 18 | 35   | 6.1 | 927    | 18 | T51317 | HSV glycoprotein D |
| c | 19 | 35   | 6.1 | 1185   | 20 | Z10959 | HSV glycoprotein D |
| c | 20 | 35   | 6.1 | 1550   | 18 | T51318 | HSV glycoprotein D |
| c | 21 | 35   | 6.1 | 1608   | 5  | N40070 | Sequence coding fo |
| c | 22 | 35   | 6.1 | 1608   | 7  | N60545 | Sequence of a port |
| c | 23 | 35   | 6.1 | 1594   | 6  | N50489 | Sequence of herpes |
| c | 24 | 34.8 | 6.1 | 1602   | 10 | N90524 | Herpes Simplex vir |
| c | 25 | 34.6 | 6.1 | 985    | 21 | A64363 | Mesothelin related |
| c | 26 | 34.6 | 6.1 | 1198   | 21 | A64362 | Mesothelin related |
| c | 27 | 34.6 | 6.1 | 1207   | 21 | A64366 | Mesothelin related |
| c | 28 | 34.6 | 6.1 | 2129   | 15 | O63972 | Soluble mesothelin |
| c | 29 | 34.6 | 6.1 | 2138   | 18 | T91079 | Megakaryocyte pote |
| c | 30 | 34.2 | 6.0 | 4308   | 20 | Z34327 | Human CAK1 antigen |
| c | 31 | 34.2 | 6.0 | 4308   | 21 | C78593 | Human EST DNA34415 |
| c | 32 | 33.8 | 5.9 | 1203   | 21 | A59324 | Coding region of a |
| c | 33 | 33.8 | 5.9 | 1632   | 19 | V68059 | Neurodegenerative  |
| c | 34 | 33.8 | 5.9 | 2516   | 21 | A59323 | DNA encoding a hum |
| c | 35 | 33.8 | 5.9 | 2711   | 19 | V68056 | Neurodegenerative  |
| c | 36 | 33.4 | 5.9 | 694    | 21 | A42007 | Human secreted exp |
| c | 37 | 33.4 | 5.9 | 1218   | 12 | Q11073 | HSV-1 antigen/heat |
| c | 38 | 33.4 | 5.9 | 1308   | 12 | Q10209 | Fusion gene compri |
| c | 39 | 33.4 | 5.9 | 1308   | 12 | Q11074 | HSV-1 antigen/heat |
| c | 40 | 33.4 | 5.9 | 1459   | 11 | Q05677 | Sequence encoding  |
| c | 41 | 33.4 | 5.9 | 2925   | 12 | Q14455 | HSV surface antige |
| c | 42 | 33   | 5.8 | 1051   | 19 | V27386 | Streptococcus pneu |
| c | 43 | 33   | 5.8 | 1319   | 18 | X30940 | Streptococcus pneu |
| c | 44 | 33   | 5.8 | 5052   | 21 | C50830 | Arabidopsis thalia |
| c | 45 | 33   | 5.8 | 13206  | 19 | V52166 | Streptococcus pneu |

## ALIGNMENTS

RESULT 1  
Z08865  
ID Z08865 standard; cDNA; 570 BP.  
XX  
AC Z08865;  
XX  
DT 04-NOV-1999 (first entry)  
XX  
DE Human interleukin B30 encoding cDNA.  
XX  
KW DNAX soluble receptor subunit 1; DNAX cytokine receptor subunit 1;  
KW interleukin B30; DSR1; DCRS1; IL-B30; cytokine receptor; diagnosis;  
KW inflammatory disorder; inflammatory response; innate immunity;  
KW morphogenic development; immunological disorder; ss.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..570  
FT FT /\*tag= a  
FT FT /product= "IL-B30"  
FT FT 1..63  
FT FT /\*tag= b  
FT FT mat\_peptide 64..567  
FT FT /\*tag= c  
XX  
XX WO9940195-A1.  
XX  
XX 12-AUG-1999.  
XX  
XX 05-FEB-1999; 99WO-US02600.  
XX  
XX 13-MAY-1998; 98US-0078194.  
XX  
XX 06-FEB-1998; 98US-0073941.  
PR





QY 421 CACTGGGAGACTCAGCAGATTCAAGCTCAGTCCAGCCAGCCATGGCAGCTCTCTTT 480  
 Db 421 cactgggagactcagcagattccaagctcagtcagtcagccagccagccagtcagtccttt 480  
 QY 481 CTCCGCTTCAAAATCTTCGAGCTCCAGCCCTTGTGGCTGTAGCCGCCGGGCTTTT 540  
 Db 481 ctccgcttcaaaatcttcgagctccagcccttgtggtgtgagccgcccgggtcttt 540  
 QY 541 GCCCATGGAGCAGCAACCTTGAGTCCCTAA 570  
 Db 541 gcccatggagcagcaaccctgagtcctaa 570

RESULT 4  
 ID 237262 standard; DNA; 1026 BP.  
 XX  
 AC 237262;  
 DT 01-FEB-2000 (first entry)  
 XX  
 DE SGRF coding sequence.  
 XX  
 KW SGRF; human; Interleukin-6 G-CSF related factor; cell proliferation;  
 KW immune system; haematopoietic system; therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO954357-A1.  
 XX  
 PD 28-OCT-1999.  
 XX  
 PF 14-APR-1999; 99WO-JP01997.  
 XX  
 PR 14-APR-1998; 98JP-0121805.  
 XX  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX  
 PI Hirata Y;  
 XX  
 DR WPI; 2000-013230/01.  
 DR P-PSDB; Y54606.  
 XX  
 PT Novel cytokine-like protein, with activity of supporting proliferation  
 PT of myeloid cells, useful in treating abnormality of cell proliferation  
 PT in immune and haematopoiesis systems  
 XX  
 PS Claim 2; Fig 1; 69pp; Japanese.  
 XX  
 CC This sequence encodes the Interleukin-6 G-CSF related factor (SGRF)  
 CC protein of the invention. The protein is a member of the IL-6/G-CSF/MSF  
 CC family. The protein can be used in drugs for treating diseases due to  
 CC abnormality of cell proliferation in the immune system and haematopoietic  
 CC system.  
 XX  
 SQ Sequence 1026 BP; 268 A; 249 C; 272 G; 237 T; 0 other;

Query Match 99.4%; Score 566.8; DB 21; Length 1026;  
 Best Local Similarity 99.6%; Pred. No. 2.2e-156;  
 Matches 568; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCTGGGGAGCAGAGCTGTAATGCTGCTGCTGCCCTGGAGCAGCTCAGGGCAGA 60  
 Db 144 atctggggagcagagctgtaatgctgctgctgccctggagcagctcagggcaga 203  
 QY 61 GCTGTGCTGGGGCAGCAGCCCTGCTGGACTCAGTCCAGCAGCTTTTACAGAAAGCTC 120  
 Db 204 gctgtgctggggcagcagccctgctggactcagtcagtcagctttcacagaagctc 263  
 QY 121 TGACACTGCTGGAGTGCACATCCACTAGTGGGACACATGATCTAAGAGAGAGGGA 180  
 Db 264 tgcacactgctggagtgcacatccactagtgggacacatggtatctaaagagaagagga 323

QY 181 GATGAAGAGAGACTAAATGATGTTCCCATATCCAGTGTGGAGATGGCTGTGACCCCAA 240  
 Db 324 gatgaagagagactcaaatgatgttcccatatccagtggtggagatggctgtgaccccaa 383  
 QY 241 GGACTCAGGGACAAACAGTCTAGTTCTGCTTGCAGAGGATCCACAGGGTCTGATTTTAT 300  
 Db 384 ggactcagggacaaacagtctagttctgcttgcagaggtccacagggctgattttat 443  
 QY 301 GAGAAGCTGCTAGGATCGGATATTTTCACAGGGGAGGCTTCTCTGCTCCCTGATAGCCCT 360  
 Db 444 gagaagctgctaggatcggaattttcacaggggagccttctctgctccctgatagacct 503  
 QY 361 GTGGCGCAGCTTCATGCTCCCTACTTGGGCTCAGCAACTCTCAGCAGCTGAGGGTCAC 420  
 Db 504 gtggcgagcttcattgctccctctactg99cctcagcagcaactcctcagcctg99gggtcac 563  
 QY 421 CACTGGGAGACTCAGCAGATTCAGCAGCTCAGTCCAGCAGCCATGGCAGCTCTCTTT 480  
 Db 564 cactgggagactcagcagattccaagcctcagtcacagcagccatggcagcgtctctt 623  
 QY 481 CTCCGCTTCAAAATCTTCGAGCTCCAGCCCTTGTGGCTGTAGCCGCCGGGCTTTT 540  
 Db 624 ctccgcttcaaaatcttcgagcctccagcctccagccttctggtgtgagccgcccgggtctt 683  
 QY 541 GCCCATGGAGCAGCAACCTTGAGTCCCTAA 570  
 Db 684 gcccatggagcagcaaccctgagtcctaa 713

RESULT 5  
 ID C84306 standard; cDNA; 1055 BP.  
 XX  
 AC C84306;  
 XX  
 DT 19-MAR-2001 (first entry)  
 XX  
 DE Human EXCS encoding cDNA (clone ID 2933038CBI).  
 XX  
 KW Extracellular signaling molecule; EXCS; anti-inflammatory; human;  
 KW immunosuppressive; cytosolic; neuroprotective; gastrointestinal;  
 KW virucide; antibacterial; anti-HIV; human immunodeficiency virus;  
 KW antiinfectility; cerebroprotective; nootropic; antitumor; antifungal;  
 KW anticonvulsant; tranquilizer; neuroleptic; vasotropic; gynecological;  
 KW keratolytic; protozoacide; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200070049-A2.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PF 19-MAY-2000; 2000WO-US13975.  
 XX  
 PR 19-MAY-1999; 99US-0134949.  
 PR 15-JUL-1999; 99US-0144270.  
 PR 30-JUL-1999; 99US-0146700.  
 PR 04-OCT-1999; 99US-0157508.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Tang YT, Yue H, Lal P, Burford N, Bandman O, Baughn MR;  
 PI Azimzai Y, Lu DAM, Patterson C;  
 XX  
 DR WPI; 2001-025021/03.  
 DR P-PSDB; B48070.  
 XX  
 PT New human extracellular signaling nucleic acids and polypeptides useful  
 PT for diagnosing, treating and preventing infections and  
 PT gastrointestinal, neurological, reproductive, and  
 PT autoimmune/inflammatory disorders -  
 XX













XX 19-FEB-2001 (first entry)  
 XX Human kinesin-like protein HKLP coding sequence contig SEQ ID NO: 1.  
 DE  
 XX  
 DE Human; kinesin-like protein; HKLP; KIF1; cell division; cancer;  
 NW intracellular transport; neurological disorder; infertility;  
 KW biallelic marker; spontaneous abortion; neonatal chromosome disorder;  
 KW aneuploidy; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200063375-A1.  
 PN  
 XX  
 XX 26-OCT-2000.  
 PD  
 XX  
 XX 20-APR-2000; 2000WO-IB00562.  
 PF  
 XX  
 XX 20-APR-1999; 99US-0130217.  
 PR  
 XX  
 XX (GEST ) GENSET.  
 PA  
 XX Bougueleret L, Dufaure-Gare I, Grel P;  
 PI  
 XX WPI: 2000-665242/64.  
 DR  
 XX  
 XX An isolated or purified human kinesin-like protein (HKLP) encoding  
 CC polynucleotide used to detect HKLP polynucleotides in a sample  
 PT comprises a contiguous span of at least 12 nucleotides -  
 PT  
 XX  
 PS Claim 1; Page 143-175; 199pp; English.  
 DR  
 XX  
 CC The present invention describes the coding and protein sequences of the  
 CC human kinesin-like protein HKLP. It is thought that the protein could be  
 CC involved in neurological disorders, infertility, spontaneous abortion, its  
 CC neonatal chromosome disorders, aneuploidy and cancers. This is due to its  
 CC function in the movement of microtubules. The protein shows homology to  
 CC the murine KIF1A and KIF1B proteins. The sequences disclosed in the  
 CC invention can be used in the isolation of similar human proteins and in  
 CC vector production. In addition, the biallelic markers shown can be used  
 CC in disease diagnosis and population studies.  
 CC  
 XX Sequence 121162 BP; 33272 A; 24108 C; 25842 G; 37919 T; 21 other;  
 SQ  
  
 Query Match 6.7%; Score 38; DB 21; Length 121162;  
 Best Local Similarity 54.2%; Pred. No. 1-5;  
 Matches 77; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
  
 QY 190 ACTACAATGATGTTCCCATATCCAGTGGAGATGGCTGTGACCCCAAGGACTCAGG 249  
 DB 87799 AATAGCAATGATTTCACTGTGAATAATGATTTATATGCTCTCTCCCCACAGCAAGC 87740  
  
 QY 250 GACACACAGTCAGTCTCTGTCGCAAGGATCCACCAGGCTGATTTTATGAGAACTG 309  
 DB 87739 GTTTTCCACAGGCTCGACACACAGGAATATCTGGAAGAATTTGTTAAACAATGGG 87680  
  
 QY 310 CTAGGATCGGATATTTTCACAG 331  
 DB 87679 CTATATTAGTATTTCACAG 87658  
  
 RESULT 13  
 ID 255567  
 XX 255567 standard; cDNA; 567 BP.  
 AC 255567;  
 XX  
 XX 14-MAR-2000 (first entry)  
 DE Feline interferon-alpha (IFN-alpha) clone #2 cDNA.  
 XX Interferon-alpha; IFN-alpha; antibody; feline; inhibitor;  
 KW

KW immune response; immunoregulation; tumour; cancer; autoimmune disease;  
 XX vaccine; ss.  
 OS  
 XX Felis catus.  
 OS  
 XX WO9961618-A2.  
 PN  
 XX 02-DEC-1999.  
 PD  
 XX  
 XX 28-MAY-1999; 99WO-US11942.  
 PF  
 XX  
 XX 29-MAY-1998; 98US-0087306.  
 PR  
 XX  
 XX (HESK-) HESKA CORP.  
 PA  
 XX  
 XX Sim G, Yang S, Dreitz MJ, Wonderling RS;  
 PI  
 XX WPI: 2000-072623/06.  
 DR  
 XX P-PSDB; Y58225.  
 DR  
 XX  
 XX Nucleic acids encoding immunoregulatory proteins from cats or dogs,  
 CC useful for treating or preventing e.g. tumors or autoimmune disease  
 PT  
 XX  
 XX Claim 1j; Page 241-242; 264pp; English.  
 PS  
 XX  
 CC Sequences 255567-255568, 255571-255572, 255569-255570 and  
 CC 255573-255574 represent cDNA sequences encoding feline encoded  
 CC and mature interferon-alpha (IFN-alpha) clones #2  
 CC and #3 respectively. The invention relates to canine  
 CC interleukin-4 (IL-4), canine or feline Flt-3 ligand, canine  
 CC or feline CD40, canine or feline CD154 (CD40 ligand), canine IL-5,  
 CC canine IL-13, feline IFN-alpha and feline granulocyte macrophage  
 CC colony-stimulating factor (GM-CSF), and nucleotides which encode these  
 CC immunoregulatory proteins. The proteins, their associated  
 CC nucleic acids, specific antibodies and inhibitors may be used as  
 CC vaccines for therapeutic or prophylactic regulation of an immune  
 CC response in animals (particularly cats, dogs, horses and humans).  
 CC They may be used to treat autoimmune or infectious diseases including  
 CC allergies, tumours, inflammation and graft rejection, and to increase  
 CC the response from a co-administered antigen. The nucleotide sequences  
 CC can also be used for the recombinant production of a protein, while  
 CC nucleotide fragments are useful as probes, as amplification primers and  
 CC as sources of inhibitory therapeutics (e.g., antisense oligonucleotides).  
 CC The proteins may be used to raise antibodies and to screen for  
 CC modulators of activity, while the antibodies may be used in detection,  
 CC and in drug targeting.  
 CC  
 XX Sequence 567 BP; 126 A; 167 C; 156 G; 118 T; 0 other;  
 SQ  
  
 Query Match 6.1%; Score 35; DB 21; Length 567;  
 Best Local Similarity 54.2%; Pred. No. 1;  
 Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
  
 QY 32 TGCTGTGCTGCCCTGGACAGCTCAGGGCAGAGCTGTGCTGGGGCAGCAGCCCTCCCTGGA 91  
 DB 2 tggcgctgcctctctcttcttggcgccctggcgctggcgctggcgcaactccgctctgct 61  
  
 QY 92 CTCAGTGCACAGCTTTTCACAGAAGCTCTGCACACTGGCTGGAGTGCACATCCACTAG 151  
 DB 62 ctctgggctgtgacctgcctccacacccacgctgtgaacaggagggccttgacgctcc 121  
  
 QY 152 TGGGACACATG 162  
 DB 122 tgggacacaaatg 132  
  
 RESULT 14  
 ID 255568/c  
 XX 255568 standard; cDNA; 567 BP.  
 AC 255568;  
 XX

DT 14-MAR-2000 (first entry)  
 XX Feline interferon-alpha (IFN-alpha) clone #2 cDNA complement.  
 DE  
 XX Interferon-alpha; IFN-alpha; antibody; feline; inhibitor;  
 KW immune response; immunoregulation; tumour; cancer; autoimmune disease;  
 KW vaccine; ss.  
 XX  
 OS Felis catus.  
 XX  
 XX WO9961618-A2.  
 PN  
 XX 02-DEC-1999.  
 PD  
 XX 28-MAY-1999; 99WO-US11942.  
 PF  
 XX 29-MAY-1998; 98US-0087306.  
 PR  
 XX (HESK-) HESKA CORP.  
 PA  
 XX Sim G.; Yang S, Dreitz MJ, Wonderling RS;  
 PI  
 XX WPI; 2000-072623/06.  
 DR  
 XX P-PSDB; Y58225.  
 XX  
 XX Nucleic acids encoding immunoregulatory proteins from cats or dogs,  
 PT useful for treating or preventing e.g. tumors or autoimmune disease  
 XX  
 PS Claim 1j; Page 243; 264pp; English.  
 XX  
 CC Sequences 255567-255568, 255571-255572, 255569-255570 and  
 CC 255573-255574 represent cDNA sequences encoding feline encoded  
 CC and mature interferon-alpha (IFN-alpha) clones #2  
 CC and #3 respectively. The invention relates to canine  
 CC interleukin-4 (IL-4), canine or feline Fit-3 ligand, canine  
 CC canine IL-13, canine or feline CD154 (CD40 ligand), canine IL-5,  
 CC canine IL-13, feline IFN-alpha and feline granulocyte macrophage  
 CC colony-stimulating factor (GM-CSF), and nucleotides which encode these  
 CC immunoregulatory proteins. The proteins, their associated  
 CC nucleic acids, specific antibodies and inhibitors may be used as  
 CC vaccines for therapeutic or prophylactic regulation of an immune  
 CC response in animals (particularly cats, dogs, horses and humans).  
 CC They may be used to treat autoimmune or infectious diseases including  
 CC allergies, tumours, inflammation and graft rejection, and to increase  
 CC the response from a co-administered antigen. The nucleotide sequences  
 CC can also be used for the recombinant production of a protein, while  
 CC nucleotide fragments are useful as probes, as amplification primers and  
 CC as sources of inhibitory therapeutics (e.g., antisense oligonucleotides).  
 CC The proteins may be used to raise antibodies and to screen for  
 CC modulators of activity, while the antibodies may be used in detection,  
 CC and in drug targeting.  
 XX  
 XX Sequence 567 BP; 118 A; 156 C; 167 G; 126 T; 0 other;

Query Match 6.1%; Score 35; DB 21; Length 567;  
 Best Local Similarity 54.2%; Pred. No. 1;  
 Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
 QY 32 TGTGTGTCCTGGAGCTTCACAGAGCTGTGCCTGGGGGACAGCCCTGCCTGGA 91  
 DB 566 TGGCGCTGCGCTTCTCTTCTTGTGGCGCTGTGGCGCTGCAACTCGTCTGCT 507  
 QY 92 CTCAGTGCCAGCAGCTTCACAGAGCTGTGCACACTGGCCTGGAGTGCACATCCACTAG 151  
 DB 506 CTTCTGGGCTGTGACCTGCCTCAGACCCAGCGCTGTGTAACAGGAGGGCTTGACGCTCC 447  
 QY 152 TGGGACACATG 162  
 DB 446 TGGGACAAATG 436

RESULT 15

255569  
 ID Z55569 standard; cDNA; 567 BP.  
 XX  
 AC 255569;  
 XX  
 DT 14-MAR-2000 (first entry)  
 DE  
 XX Feline interferon-alpha (IFN-alpha) clone #3 cDNA.  
 XX Interferon-alpha; IFN-alpha; antibody; feline; inhibitor;  
 KW immune response; immunoregulation; tumour; cancer; autoimmune disease;  
 KW vaccine; ss.  
 XX  
 OS Felis catus.  
 XX  
 XX WO9961618-A2.  
 PN  
 XX 02-DEC-1999.  
 PD  
 XX 28-MAY-1999; 99WO-US11942.  
 PF  
 XX 29-MAY-1998; 98US-0087306.  
 PR  
 XX (HESK-) HESKA CORP.  
 PA  
 XX Sim G, Yang S, Dreitz MJ, Wonderling RS;  
 PI  
 XX WPI; 2000-072623/06.  
 DR  
 XX P-PSDB; Y58226.  
 XX  
 XX Nucleic acids encoding immunoregulatory proteins from cats or dogs,  
 PT useful for treating or preventing e.g. tumors or autoimmune disease  
 XX  
 PS Claim 1j; Page 244-245; 264pp; English.  
 XX  
 CC Sequences 255567-255568, 255571-255572, 255569-255570 and  
 CC 255573-255574 represent cDNA sequences encoding feline encoded  
 CC and mature interferon-alpha (IFN-alpha) clones #2  
 CC and #3 respectively. The invention relates to canine  
 CC interleukin-4 (IL-4), canine or feline Fit-3 ligand, canine  
 CC canine IL-13, canine or feline CD154 (CD40 ligand), canine IL-5,  
 CC canine IL-13, feline IFN-alpha and feline granulocyte macrophage  
 CC colony-stimulating factor (GM-CSF), and nucleotides which encode these  
 CC immunoregulatory proteins. The proteins, their associated  
 CC nucleic acids, specific antibodies and inhibitors may be used as  
 CC vaccines for therapeutic or prophylactic regulation of an immune  
 CC response in animals (particularly cats, dogs, horses and humans).  
 CC They may be used to treat autoimmune or infectious diseases including  
 CC allergies, tumours, inflammation and graft rejection, and to increase  
 CC the response from a co-administered antigen. The nucleotide sequences  
 CC can also be used for the recombinant production of a protein, while  
 CC nucleotide fragments are useful as probes, as amplification primers and  
 CC as sources of inhibitory therapeutics (e.g., antisense oligonucleotides).  
 CC The proteins may be used to raise antibodies and to screen for  
 CC modulators of activity, while the antibodies may be used in detection,  
 CC and in drug targeting.  
 XX  
 XX Sequence 567 BP; 123 A; 169 C; 158 G; 117 T; 0 other;

Query Match 6.1%; Score 35; DB 21; Length 567;  
 Best Local Similarity 54.2%; Pred. No. 1;  
 Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
 QY 32 TGTGTGTCCTGGAGCTTCACAGAGCTGTGCCTGGGGGACAGCCCTGCCTGGA 91  
 DB 2 TGGCGCTGCGCTTCTCTTCTTGTGGCGCTGTGGCGCTGCAACTCGTCTGCT 61  
 QY 92 CTCAGTGCCAGCAGCTTCACAGAGCTTCACAGAGCTGTGCCTGGAGTGCACATCCACTAG 151  
 DB 62 cctctgggctgtgacctgctcagaccacggcctgctgaacaggaggccttgacctcc 121  
 QY 152 TGGGACACATG 162

Db 122 tgggacaaatg 132

Search completed: June 21, 2001, 02:08:54  
Job time: 4583 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2001, 23:16:50 ; Search time 1723.79 Seconds  
(without alignments)  
3125.744 Million cell updates/sec

Title: US-09-558-474-1  
Perfect score: 570  
Sequence: 1 ATGCTGGGGAGCAGAGCTGT.....CAGCACCCCTGAGTCCCTAA 570

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estrol:\*  
83: em\_estrol2:\*  
84: em\_estrol3:\*  
85: em\_estrol4:\*  
86: em\_estrol5:\*  
87: em\_estrol6:\*  
88: em\_estrol7:\*  
89: em\_estrol8:\*  
90: em\_estrol9:\*  
91: em\_estrol10:\*  
92: em\_estrol11:\*  
93: em\_estrol12:\*  
94: em\_estrol13:\*  
95: em\_estrol14:\*  
96: em\_estrol15:\*  
97: em\_estrol16:\*  
98: em\_estrol17:\*  
99: em\_estrol18:\*  
100: em\_estrol19:\*  
101: em\_estrol20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



SUMMARIES

| Result No. | Score | Query Match | Length | DB  | ID       | Description |
|------------|-------|-------------|--------|-----|----------|-------------|
| 1          | 404.8 | 71.0        | 833    | 104 | AJ277049 |             |
| 2          | 393   | 68.9        | 492    | 7   | AA418955 |             |
| 3          | 321.2 | 56.4        | 1150   | 137 | BE572471 |             |
| 4          | 232.8 | 40.8        | 542    | 102 | AI796983 |             |
| 5          | 231.4 | 40.6        | 897    | 137 | BE571481 |             |
| 6          | 220   | 38.6        | 372    | 155 | C06368   |             |
| 7          | 210.4 | 36.9        | 430    | 141 | BE847383 |             |
| 8          | 206.8 | 36.3        | 719    | 153 | BE435629 |             |
| 9          | 183   | 32.1        | 496    | 175 | BE236314 |             |
| 10         | 150   | 26.3        | 457    | 117 | AW511070 |             |
| 11         | 140.2 | 24.6        | 367    | 145 | BF189762 |             |
| 12         | 138.6 | 24.3        | 575    | 31  | AV598349 |             |
| 13         | 122.8 | 21.5        | 399    | 114 | AW316433 |             |
| 14         | 113.2 | 19.9        | 566    | 165 | BE237060 |             |
| 15         | 106.4 | 18.7        | 323    | 129 | BB270313 |             |
| 16         | 102   | 17.9        | 320    | 130 | BB274745 |             |
| 17         | 91.6  | 16.1        | 137    | 188 | TI8577   |             |
| 18         | 90    | 15.8        | 400    | 14  | AI004049 |             |
| 19         | 58    | 10.2        | 369    | 111 | AW072323 |             |
| 20         | 54    | 9.5         | 397    | 7   | AA418747 |             |
| 21         | 39.4  | 6.9         | 404    | 170 | BF834812 |             |
| 22         | 38.8  | 6.8         | 579    | 141 | BE884343 |             |
| 23         | 38.6  | 6.8         | 1101   | 219 | CNS00LXJ |             |
| 24         | 38.4  | 6.7         | 198    | 11  | AA738190 |             |
| 25         | 38    | 6.7         | 558    | 140 | BE809252 |             |
| 26         | 37.8  | 6.6         | 459    | 234 | AO852080 |             |
| 27         | 37.4  | 6.6         | 889    | 145 | BF204344 |             |
| 28         | 37.2  | 6.5         | 383    | 7   | AA415467 |             |
| 29         | 37    | 6.5         | 421    | 12  | AA794290 |             |
| 30         | 37    | 6.5         | 539    | 220 | CNS02760 |             |
| 31         | 37    | 6.5         | 600    | 156 | C77322   |             |
| 32         | 36.8  | 6.5         | 355    | 142 | BE935507 |             |
| 33         | 36.8  | 6.5         | 440    | 7   | AA434734 |             |
| 34         | 36.8  | 6.5         | 522    | 136 | BE494337 |             |
| 35         | 36.8  | 6.5         | 831    | 220 | CNS01YR4 |             |
| 36         | 36.6  | 6.4         | 376    | 2   | AA110023 |             |
| 37         | 36.4  | 6.4         | 609    | 175 | BG300864 |             |
| 38         | 36.4  | 6.4         | 646    | 143 | BF043357 |             |
| 39         | 36.2  | 6.4         | 409    | 138 | BE655580 |             |
| 40         | 36    | 6.3         | 324    | 167 | BE400665 |             |
| 41         | 36    | 6.3         | 469    | 121 | AW839214 |             |
| 42         | 36    | 6.3         | 933    | 106 | AL578453 |             |
| 43         | 36    | 6.3         | 989    | 106 | AL578994 |             |
| 44         | 36    | 6.3         | 1037   | 106 | AL574127 |             |
| 45         | 35.8  | 6.3         | 469    | 16  | AI120671 |             |

RESULT 1  
AJ277049  
LOCUS  
DEFINITION AJ277049 Placenta Matchmaker cDNA library(Clontech) Homo sapiens  
ACCESSION AJ277049  
VERSION AJ277049.1 GI:7342435  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 833)  
Bach,S., Portetelle,D. and Vandenbol,M.  
Isolation of new human EST sequences by using the MSB4 yeast gene as bait in the two-hybrid screening of a Human Placenta cDNA library

JOURNAL COMMENT

Unpublished (2000)  
Contact: Bach SL  
Microbiology Unit  
Gembloux Agricultural University  
Avenue Marechal Juin 6, B-5030 Gembloux, Belgium.  
Location/Qualifiers

FEATURES

Source

1. 833  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="BHA2"  
/clone\_lib="Placenta Matchmaker cDNA library(Clontech)"  
/sex="Female"  
/tissue\_type="Placenta"

BASE COUNT 227 a 197 c 200 g 209 t  
ORIGIN

Query Match 71.0%; Score 404.8; DB 104; Length 833;  
Best Local Similarity 99.5%; Pred. No. 1.2e-97;  
Matches 406; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 163 GATCTAAGAGAGAGGAGGATGAAGAGACTACAAATGATGTTCCCATATCCAGTGTGA 222  
DB 97 GATCTAAGAGAGAGGAGGATGAAGAGACTACAAATGATGTTCCCATATCCAGTGTGA 156  
QY 223 GATGGTGTGACCCCAAGGACTCAGGGACACAGTCAGTCTTCTGCTTGCAGAGATCCAC 282  
DB 157 GATGGTGTGACCCCAAGGACTCAGGGACACAGTCAGTCTTCTGCTTGCAGAGATCCAC 216  
QY 283 CAGGGTCTGATTTTATGAGAAGCTGCTAGGATCGGATATTTTACAGGGAGCCCTTCT 342  
DB 217 CAGGGTCTGATTTTATGAGAAGCTGCTAGGATCGGATATTTTACAGGGAGCCCTTCT 276  
QY 343 CTGCTCCCTGATAGCCCTGTGGGAGCTTCATGCTCCCTACTGGGCTCAGCCCAACTC 402  
DB 277 CTGCTCCCTGATAGCCCTGTGGGAGCTTCATGCTCCCTACTGGGCTCAGCCCAACTC 336  
QY 403 CTGACGCTGAGGTGACCACTGGGAGCTCAGCAGATCCAGCCTCAGTCCAGCCAG 462  
DB 337 CTGACGCTGAGGTGACCACTGGGAGCTCAGCAGATCCAGCCTCAGTCCAGCCAG 396  
QY 463 CCATGCGAGCGTCTCTCTTCCGCTTCAAAATCTTCGAGCCTCCAGGCTTTGTGGCT 522  
DB 397 CCATGCGAGCGTCTCTCTTCCGCTTCAAAATCTTCGAGCCTCCAGGCTTTGTGGCT 456  
QY 523 GTAGCCGCCGGTCTTTGCCCATGAGCAGCAACCTCAGTCCCTAA 570  
DB 457 GTAGCCGCCGGTCTTTGCCCATGAGCAGCAACCTCAGTCCCTAA 504

RESULT 2  
AA418955

LOCUS  
DEFINITION zw01c10.r1 Soares\_NhMPu\_S1 Homo sapiens cDNA clone IMAGE:768018  
5', mRNA sequence.

ACCESSION AA418955  
VERSION AA418955.1 GI:2080765  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 492)  
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Willson,R.

WashU-Merck EST Project 1997  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800

TITLE  
JOURNAL  
COMMENT

ALIGNMENTS

AA418955 492 bp mRNA EST 16-OCT-1997  
zw01c10.r1 Soares\_NhMPu\_S1 Homo sapiens cDNA clone IMAGE:768018  
5', mRNA sequence.  
AA418955  
AA418955.1 GI:2080765  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 492)  
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Willson,R.  
WashU-Merck EST Project 1997  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800

Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -28ml3 rev2 ET from Amersham.

[illegible]

|            |       |       |       |       |
|------------|-------|-------|-------|-------|
| BASE COUNT | 117 a | 136 c | 136 q | 103 t |
|------------|-------|-------|-------|-------|

|                       |         |                    |       |             |
|-----------------------|---------|--------------------|-------|-------------|
| Query Match           | 68.9%;  | Score 393;         | DB 7; | Length 492; |
| Best Local Similarity | 100.0%; | Pred. No. 1.5e-94; |       |             |

|     |    |   |     |
|-----|----|---|-----|
| 163 | QY | GATCTAAGAGAGAGGGAGATGAAGAGACTACAATGATGTTCGCCATATCCAGTGTGGA      | 222 |
|     |    |   |     |
|     |    |   |     |
| 100 | Db | GATCTAAGAGAGAGGGAGATGAAGAGACTACAATGATGTTCGCCATATCCAGTGTGGA      | 159 |
|     |    |   |     |
|     |    |   |     |
| 223 | QY | GATGGCTGTGACCCCAAGAGACTCAGGGACAACAGTCAGTTCGCTTGCAAAAGATCCAC     | 282 |
|     |    |   |     |
|     |    |   |     |
| 160 | Db | GATGGCTGTGACCCCAAGAGACTCAGGGACAACAGTCAGTTCGCTTGCAAAAGATCCAC     | 219 |
|     |    |   |     |
|     |    |   |     |
| 283 | QY | CAGGGCTGTGATTTTTATGAGAAGCTCTAGGATCGGATATTTTACAGGGGAGCCTTCT      | 342 |
|     |    |   |     |
|     |    |   |     |
| 220 | Db | CAGGGCTGTGATTTTTATGAGAAGCTCTAGGATCGGATATTTTACAGGGGAGCCTTCT      | 279 |
|     |    |   |     |
|     |    |   |     |
| 343 | QY | CTGCTCCCTCATAGCCCTGTGGCGCAGCTTCATGCCCTCCTACTGGGCGTCAGGCCAATC    | 402 |
|     |    |   |     |
|     |    |   |     |
| 280 | Db | CTGCTCCCTCATAGCCCTGTGGCGCAGCTTCATGCCCTCCTACTGGGCGTCAGGCCAATC    | 339 |
|     |    |   |     |
|     |    |   |     |
| 403 | QY | CTGCAGGCTCAGGGGTCAACCACTGGGAGACCTCAGCAGATTCCAAGCCCTCAGTCCAGCCAG | 462 |
|     |    |   |     |
|     |    |   |     |
| 340 | Db | CTGCAGGCTCAGGGGTCAACCACTGGGAGACCTCAGCAGATTCCAAGCCCTCAGTCCAGCCAG | 399 |
|     |    |   |     |
|     |    |   |     |
| 463 | QY | CCATGGCAGGGTCTCTTCTCCGGTTCAAAATCCTTCGCAGCCCTCCAGGCGTTTGTGGCT    | 522 |
|     |    |   |     |
|     |    |   |     |
| 400 | Db | CCATGGCAGGGTCTCTTCTCCGGTTCAAAATCCTTCGCAGCCCTCCAGGCGTTTGTGGCT    | 459 |
|     |    |   |     |
|     |    |   |     |
| 523 | QY | GTAGCCGCCGGGTCTTTTGCCCATGGAGCAGCA                               | 555 |
|     |    |   |     |
|     |    |   |     |
| 460 | Db | GTAGCCGCCGGGTCTTTTGCCCATGGAGCAGCA                               | 492 |
|     |    |   |     |
|     |    |   |     |

|            |  |            |
|------------|--|------------|
| RESULT     | 3  |            |
| BE572471   |  |            |
| LOCUS      |  |            |
| DEFINITION | BE572471 1150 bp mRNA EST 15-AUG-2000<br>601331259F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3708687 5', |            |
| ACCESSION  | BE572471   |            |
| VERSION    | BE572471.1   | GI:9816191 |
| KEYWORDS   | EST.   |            |
| SOURCE     | house mouse.   |            |
| ORGANISM   | Mus musculus   |            |
|            | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  |            |

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1150)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs@email.nih.gov](mailto:cgapbs@email.nih.gov)  
 Tissue procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:

## FEATURES

|  |            |
|--|------------|
|  | BASE COUNT |
|  | ORIGIN     |

Query Match  
Best Local  
Matches 4

|     |    |     |    |
|-----|----|-----|----|
| 1   | AY | 1   | AY |
| 78  | AY | 78  | AY |
| 60  | AY | 60  | AY |
| 138 | AY | 138 | AY |
| 118 | AY | 118 | AY |
| 198 | AY | 198 | AY |
| 173 | AY | 173 | AY |
| 258 | AY | 258 | AY |
| 233 | AY | 233 | AY |
| 318 | AY | 318 | AY |
| 293 | AY | 293 | AY |
| 378 | AY | 378 | AY |
| 352 | AY | 352 | AY |
| 438 | AY | 438 | AY |
| 412 | AY | 412 | AY |
| 498 | AY | 498 | AY |
| 471 | AY | 471 | AY |

```
Db 558 GCGCCCTCTCCGTTCCAGATCTCTCGAAGCCTCCAGGCCTTTTGGCCATAGCTGC 617
QY 531 CCGGGTCTTGGCCATGGAGCAGCAACCTTGACT 564
Db 618 CCGGGTCTTGGCCAGGAGCAGCAACTCTGACT 651

RESULT 4
LOCUS AI796983/c 542 bp mRNA EST 18-DEC-1999
DEFINITION we25e07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2342148 3',
mRNA sequence.
ACCESSION AI796983
VERSION AI796983.1 GI:5362455
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 542)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Cloning Distribution: Washington University Genome Sequencing Center
Cloning Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbbrp/image/image.html
Insert Length: 868 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 449.

FEATURES
Location/Qualifiers
1..542
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2342148"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="Carcinoid"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonoids
141920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 150 a 125 c 140 g 127 t
ORIGIN
Query Match 40.8%; Score 232.8; DB 102; Length 542;
Best Local Similarity 99.2%; Pred. No. 1e-51;
Matches 234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 335 AGCCTTCTGCTCCCTGATAGCCCTGTGGCGAGCTTCATGCTCCCTACTGGGCTCA 394
Db 542 AGCCTTCTGCTCCCTGATAGCCCTGTGGCGAGCTTCATGCTCCCTACTGGGCTCA 483
QY 395 GCCAATCTCTGAGCCTGAGGGTACCACCTGGGAGACTCAGCAGATTCGAAGCCTCAGTC 454
Db 482 GCCAATCTCTGAGCCTGAGGGTACCACCTGGGAGACTCAGCAGATTCGAAGCCTCAGTC 423
QY 455 CCAGCAGCAGCATGGCAGCGTCTCTTCTTCGCGTTCAAAATCCCTTCGAGCCTCCAGGCCT 514
Db 422 CCAGCAGCAGCATGGCAGCGTCTCTTCTTCGCGTTCAAAATCCCTTCGAGCCTCCAGGCCT 363
```

```
QY 515 TTGTGGCTGTAGCCGCCCGGGTCTTTGCCATGGAGCAACCCCTGAGTCCCTAA 570
Db 362 TTGTGGCTGTAGCCGCCCGGGTCTTTGCCATGGAGCAACCCCTGAGTCCCTAA 307

RESULT 5
LOCUS BE571481 897 bp mRNA EST 15-AUG-2000.
DEFINITION 601333907F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3711529 5',
mRNA sequence.
ACCESSION BE571481
VERSION BE571481
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 897)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Cloning Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8935 row: m column: 02
High quality sequence stop: 597.

FEATURES
Location/Qualifiers
1..897
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3711529"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Salt;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo df.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 239 a 201 c 262 g 194 t
ORIGIN
Query Match 40.6%; Score 231.4; DB 137; Length 897;
Best Local Similarity 73.4%; Pred. No. 2.8e-51;
Matches 391; Conservative 0; Mismatches 131; Indels 11; Gaps 7;

QY 7 GGCAGCAGAGCTGTAATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 66
Db 117 GGATGGAGAGCAGTAATAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 175
QY 67 CCTGGGGCAGCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 126
Db 176 CCTAGGAGTAGCAGTCTGAGTGGGCTCAGTGGCAGCAGCTCTCTCGAATCTCTG 235
QY 127 CTGCGCTGGAGTGCACATCCACCTAGTGGGACACATGGAT---CTAAGAGAAGAGGAGAT 183
Db 236 CTAGCCTGGAACGCACATGCACCGGGGACATATGAATCTACTAAGAGAAGAGAGAT 295
QY 184 GAGAGACTACAATATGATGTTTCCCATATCCAGTGGGAGATGGCTGTGACCCCAAGGA 243
Db 296 GAGAGACTAAAAAATATGTGCCCCGTATCCAGTGTGAAGATGTTGTGTGACCCCAAGGA 355
QY 244 CTCAGGAGACAACAGTCTGCTGCTTGC- AAAGGATCCACAGGCTGCTGATTTTTTATGA 302
```

|                       |  |   |                  |
|-----------------------|--|---|------------------|
| Qy                    | 181  | GATGAAGAGACTACAAATGATGTCTCCCATATCCAGTGTGGGA   | 222              |
| Dd                    | 331  | GATGAAGAGACTACNAATGATGTCTCCCATATCCAGTGTGGGA   | 372              |
|                       |  |   |                  |
| RESULT                | 7  |   |                  |
| BE847383              |  |   |                  |
| LOCUS                 |  | 430 bp mRNA EST   | 26-SEP-2000      |
| DEFINITION            |  | uw26b04.yl Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:3417775 5', mRNA sequence. |                  |
| VERSION               | BE847383   |   |                  |
| KEYWORDS              | BE847383.1 GI:10305722   |   |                  |
| SOURCE                | EST.   |   |                  |
| ORGANISM              | house mouse.<br>Mus musculus   |   |                  |
| REFERENCE             | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;<br>Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus<br>1 (bases 1 to 430)   |   |                  |
| AUTHORS               | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  |   |                  |
| TITLE                 | National Cancer Institute, Cancer Genome Anatomy Project (CGAP),<br>Unpublished (1997)   |   |                  |
| JOURNAL               | Tumor Gene Index   |   |                  |
| COMMENT               | Contact: Robert Strausberg, Ph.D.<br>Email: cgapbs-re@mail.nih.gov<br>This clone is available royalty-free through LLNL ; contact the<br>IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information.<br>MG1:1093587 |   |                  |
| FEATURES              | Seq primer: -4ORP from Gibco.<br>Location/Qualifiers   |   |                  |
| Source                | 1..430<br>/organism="Mus musculus"<br>/strain="C57BL/6J"<br>/db_xref="taxon:10090"<br>/clone="IMAGE:341775"<br>/clone_lib="Soares_thymus_2NbMT"<br>/sex="male"<br>/tissue_type="Thymus"<br>/lab_stage="4 weeks"<br>/lab_host="DH10B"                               |   |                  |
|                       |  |   |                  |
| BASE COUNT            | 120 a 99 c 123 g 88 t  |   |                  |
| ORIGIN                |  |   |                  |
| Query Match           | 36.9%; Score 210.4;  | DB 141;   | Length 430;      |
| Best Local Similarity | 81.3%;   | Pred No. 9, 5e-46;  |                  |
| Matches 257;          | Conservative 0;  | Mismatches 56;  | Indels 3; Gaps 1 |
|                       |  |   |                  |
| QY                    | 1  | ATGCTGGGAGCAGAGCTAATGCTGTGTTCCTGCCCTGGACAGCTCAGGGCAGA                                   | 60               |
| Dd                    | 111  | ATGCTGGATTGCAGAGCAGTAATAATGCTATGGCTGTTCCTGGGTCACTCAGGGCCTG                              | 170              |
|                       |  |   |                  |
| QY                    | 61   | GCTGTGCCTGGGGCAGCACGCCCTGCCTGGACTCAGTGCCAGCAGCTTTTCAGANGCTC                             | 120              |
| Dd                    | 171  | GCTGTGCCTTAGGAGTAGCAGTCCCTCACTGGGCTAGTGCCAGCAGCTCTCTCGGAATCTC                           | 230              |
|                       |  |   |                  |
| QY                    | 121  | TGCACACTGGCCTGGAGTGCACATCCACTAGTGGGACACATGGAT---CTAAGAGAAGAC                            | 177              |
| Dd                    | 231  | TGCATGCTAGCCTGGAACCCACATGCACCAGCGGACATATGAATCTACTRAGAGAAGA                              | 290              |
|                       |  |   |                  |
| QY                    | 178  | GGAGATGAAGAGACTACAAATGATGTTCCTCCCATATCCAGTGTGGAGATGGGTGTGACCCC                          | 237              |
| Dd                    | 291  | GAGAGTGAAGAGACTAAAAATGATGTCCCCCTATCCAGTGTGAAGATGGTGTGACCCA                              | 350              |

```

QY 238 CAAGGACTCAGGACACACAGTCAGTCTGCTTCAAGAGTCCACCAGGCTCTGATTTT 297
|||||
Db 351 CAAGGACTCAGGACACACAGTCAGTCTGCTTCAAGAGTCCGCCAAGGCTGTGGCTTTT 410
|||||

QY 298 TATGAGAAGCTGCTAG 313
|||||
Db 411 TATAAGCACCTGCTTG 426

RESULT 8
Bg435629 719 bp mRNA EST 14-MAR-2001
LOCUS 602506911f1 NTH_MGC_79 Homo sapiens cDNA clone IMAGE:4604139 5',
DEFINITION mRNA sequence.
ACCESSION Bg435629
VERSION Bg435629.1 GI:13342135
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 719)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Inocyte Genomics, Inc.
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI339 row: e column: 04
High quality sequence stop: 496.

FEATURES
Location/Qualifiers
1..719
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4604139"
/lab_host="NIH_MGC_79"
/note="Organ: placenta; Vector: pDNR-LIB (Clontech);
Site_1: SfII (ggcgctcgcc); Site_2: SfiI (ggccattatggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3'
(where B = A, C, G, or T). Average
insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."

BASE COUNT 185 a 184 c 179 g 171 t
ORIGIN

Query Match 36.3%; Score 206.8; DB 153; Length 719;
Best Local Similarity 99.0%; Pred. No. 1e-44;
Matches 208; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 361 GTGGCGCAGCTTCATGCCCTCCCTACTGGGCTCAGCCCACTCTCGAGCCTGAGGGTCAC 420
|||||
Db 1 GTGGGCGCAGCTTCATGCCCTCCCTACTGGGCTCAGCCCACTCTCGAGCCTGAGGGTCAC 60
|||||

QY 421 CACTGGGAGACTCAGCAGATTCGAAGCCTCAGTCCCGCAGCCAGCCATGGCAGCGTCCTCCTT 480
|||||
Db 61 CACTGGGAGACTCAGCAGATTCGAAGCCTCAGTCCCGCAGCCAGCCATGGCAGCGTCCTCCTT 120
|||||

QY 481 CTCGCTTCAAAATCCTTCGACGCTCCAGGCTTTGTGGCTGTAGCCGCCCGGGGCTTTT 540
|||||
Db 121 CTCGCTTCAAAATCCTTCGACGCTCCAGGCTTTGTGGCTGTAGCCGCCCGGGGCTTTT 180
|||||

```

```

QY 541 GCCCATGGAGCAGCACACCTGAGTCCCTAA 570
|||||
Db 181 GCCCATGGAGCAGCACACCTGAGTCCCTAA 210
|||||

RESULT 9
Bg236314/c 496 bp mRNA EST 12-FEB-2001
LOCUS naf26b02.x1 Soares_NPBMC Homo sapiens cDNA clone IMAGE:4142138 3',
DEFINITION mRNA sequence.
ACCESSION Bg236314
VERSION Bg236314.1 GI:12750161
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 496)
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: M. Bento Soares, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D. and M. Fatima
Bonaldo, Ph.D.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 479.

FEATURES
Location/Qualifiers
1..496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4142138"
/lab_host="Soares_NPBMC"
/tissue_type="lymphocyte"
/note="Organ: blood; Vector: pT7T3D-Pac; Site_1: NotI;
Site_2: EcoRI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTTCACCATCTGAAGTGGGAGCGGCCGCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. Library
is normalized; constructed in the laboratory of M. Bento
Soares (University of Iowa)."

BASE COUNT 134 a 114 c 119 g 129 t
ORIGIN

Query Match 32.1%; Score 183; DB 175; Length 496;
Best Local Similarity 100.0%; Pred. No. 2.1e-38;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 GGCTTCAGCCCACTCTCGAGCCTGAGGGTCACACTGGGAGACTCAGCAGATTCCAAGC 447
|||||
Db 496 GGCTTCAGCCCACTCTCGAGCCTGAGGGTCACACTGGGAGACTCAGCAGATTCCAAGC 437
|||||

QY 448 CTCAGTCCCAGCAGCCATGGCAGCGTCTCTTCTCCGCTTCAAAATCCTTCGACGCTC 507
|||||
Db 436 CTCAGTCCCAGCAGCCATGGCAGCGTCTCTTCTCCGCTTCAAAATCCTTCGACGCTC 377
|||||

QY 508 CAGGCTTTGTGGCTGTAGCCGCCCGGGTCTTGGCCATGGAGCAGCAACCTTGAGTCCC 567
|||||
Db 376 CAGGCTTTGTGGCTGTAGCCGCCCGGGTCTTGGCCATGGAGCAGCAACCTTGAGTCCC 317
|||||

QY 568 TAA 570

```

```

Db 316 TAA 314
||||
RESULT 10
AW511070/c
LOCUS hd38e10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2911818 3', mRNA sequence.
ACCESSION AW511070
VERSION AW511070.1 GI:7149148
KEYWORDS EST.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 457)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400p from Gibco
High quality sequence stop: 439.
FEATURES
source
location/Qualifiers
1..457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2911818"
/clone="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI-CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 582632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 129 a 105 c 106 g 117 t
ORIGIN
Query Match 26.3%; Score 150; DB 117; Length 457;
Best Local Similarity 100.0%; Pred. No. 1.4e-29;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 421 CACTGGGAGACTCAGCAGATTCCAGCCTCAGTCCCGAGCCAGCCAGCCGTCCTT 480
|||||
Db 457 CACTGGGAGACTCAGCAGATTCCAGCCTCAGTCCCGAGCCAGCCAGCCGTCCTT 398
QY 481 CTCGCTTCAAAATCCCTCCAGCCTCCAGCCCTTTGCGCTGTAGCCGCCGGTCTTT 540
|||||
Db 397 CTCGCTTCAAAATCCCTCCAGCCTCCAGCCTTTGCGCTGTAGCCGCCGGTCTTT 338
QY 541 GCCCATGGAGCAGCAACCCCTGAGTCCCTAA 570
|||||
Db 337 GCCCATGGAGCAGCAACCCCTGAGTCCCTAA 308
RESULT 11
BF189762/c
LOCUS BF189762 367 bp mRNA EST
DEFINITION 235459 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF189762
VERSION BF189762.1 GI:11073131

```

---

```

KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 367)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 58 row: E column: 22
Seq primer: ATTAGGTGACACTATAG.
FEATURES
source
location/Qualifiers
1..367
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 75 a 100 c 107 g 85 t
ORIGIN
Query Match 24.6%; Score 140.2; DB 145; Length 367;
Best Local Similarity 89.3%; Pred. No. 5.3e-27;
Matches 151; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 402 CCTCAGCCTGAGGTCACCACTGGGAGACTCAGAGATTCCAAAGCTCAGTCCAGCCA 461
|||||
Db 270 CTTTCAGCCGAGGTCACCACTGGGAGACTGAGCAGACGCCAAGCCAGTCCAGCCA 211
QY 462 GCCATGGCAGCGTCTCCTTCTCCGCTTCAAAATCCTTCGAGCCTCCAGGCTTTGTGGC 521
|||||
Db 210 GCCCTGGCAACGCCCTCCTTCTCCGCTTCAAGATCCTTCGAGCCTCCAGGCTTTGTGGC 151
QY 522 TGTAGCCGCGGGTCTTTGCCCATGGAGCAGCAACCTGAGTCCCTAA 570
|||||
Db 150 TGTAGTCTGCCGGGTCTTCGCCCATGGAGCAGCAACTCTGAGCCAGTAA 102
RESULT 12
AV598349/c
LOCUS AV598349 575 bp mRNA EST
DEFINITION AV598349 Bos taurus cartilage fetus Bos taurus cDNA clone
EICA037E02 5', mRNA sequence.
ACCESSION AV598349
VERSION AV598349.1 GI:9715927
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 575)
AUTHORS Sugimoto,Y., Hirotsune,S., Takasuga,A., Itoh,R., Jitchazono,A. and
Suzuki,H.

```











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 02:07:38 ; Search time 2434.93 seconds  
(without alignments)  
7641.987 Million cell updates/sec

Title: US-09-558-474-3  
Perfect score: 1203  
Sequence: 1 CGCTAGAGTCGGACTACA.....TACCAATACGAACAGACAAA (1203)

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl:\*
- 1: gb\_ba1:\*
  - 2: gb\_ba2:\*
  - 3: gb\_ba3:\*
  - 4: gb\_in1:\*
  - 5: gb\_in2:\*
  - 6: gb\_in3:\*
  - 7: gb\_om:\*
  - 8: gb\_ov:\*
  - 9: gb\_pat1:\*
  - 10: gb\_pat2:\*
  - 11: gb\_ph:\*
  - 12: gb\_pl1:\*
  - 13: gb\_pl2:\*
  - 14: gb\_pl3:\*
  - 15: gb\_pl4:\*
  - 16: em\_ba1:\*
  - 17: em\_ba2:\*
  - 18: em\_fun:\*
  - 19: em\_htgo\_hum:\*
  - 20: em\_htgo\_inv:\*
  - 21: em\_htgo\_rod:\*
  - 22: em\_htg\_hum1:\*
  - 23: em\_htg\_hum2:\*
  - 24: em\_htg\_hum3:\*
  - 25: em\_htg\_hum4:\*
  - 26: em\_htg\_hum5:\*
  - 27: em\_htg\_hum6:\*
  - 28: em\_htg\_hum7:\*
  - 29: em\_htg\_hum8:\*
  - 30: em\_htg\_inv1:\*
  - 31: em\_htg\_inv2:\*
  - 32: em\_htg\_other:\*
  - 33: em\_htg\_rod:\*
  - 34: em\_hum1:\*
  - 35: em\_hum2:\*
  - 36: em\_hum3:\*
  - 37: em\_hum4:\*
  - 38: em\_hum5:\*
  - 39: em\_hum6:\*
  - 40: em\_hum7:\*
  - 41: em\_in:\*
  - 42: em\_om:\*
  - 43: em\_or:\*

- 44: em\_ov:\*
- 45: em\_pat:\*
- 46: em\_ph:\*
- 47: em\_pl:\*
- 48: em\_ro:\*
- 49: em\_sts:\*
- 50: em\_sy:\*
- 51: em\_un:\*
- 52: em\_vi:\*
- 53: gb\_sts1:\*
- 54: gb\_sts2:\*
- 55: gb\_sts3:\*
- 56: gb\_sy:\*
- 57: gb\_un:\*
- 58: gb\_v1:\*
- 59: gb\_v12:\*
- 60: gb\_htg1:\*
- 61: gb\_htg2:\*
- 62: gb\_htg3:\*
- 63: gb\_htg4:\*
- 64: gb\_htg5:\*
- 65: gb\_htg6:\*
- 66: gb\_htg7:\*
- 67: gb\_htg8:\*
- 68: gb\_htg9:\*
- 69: gb\_htg10:\*
- 70: gb\_htg11:\*
- 71: gb\_htg12:\*
- 72: gb\_htg13:\*
- 73: gb\_htg14:\*
- 74: gb\_htg15:\*
- 75: gb\_htg16:\*
- 76: gb\_htg17:\*
- 77: gb\_htg18:\*
- 78: gb\_htg19:\*
- 79: gb\_htg20:\*
- 80: gb\_htg21:\*
- 81: gb\_htg22:\*
- 82: gb\_htg23:\*
- 83: gb\_htg24:\*
- 84: gb\_htg25:\*
- 85: gb\_pr1:\*
- 86: gb\_pr2:\*
- 87: gb\_pr3:\*
- 88: gb\_pr4:\*
- 89: gb\_pr5:\*
- 90: gb\_pr6:\*
- 91: gb\_pr7:\*
- 92: gb\_pr8:\*
- 93: gb\_pr9:\*
- 94: gb\_rol:\*
- 95: gb\_rod:\*
- 96: gb\_in4:\*
- 97: gb\_pr10:\*
- 98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description        |
|------------|-------|---------|--------------|-------|--------------------|
| 1          | 1203  | 100.0   | 1203         | 10    | AX097623 Sequence  |
| 2          | 1203  | 100.0   | 1358         | 94    | AF301619 Mus muscu |
| 3          | 441.4 | 36.7    | 1025         | 89    | AF301620 Homo sapi |
| 4          | 439.8 | 36.6    | 1026         | 85    | AB030000 Homo sapi |
| 5          | 439.8 | 36.6    | 1055         | 9     | AX048200 Sequence  |
| 6          | 391   | 32.5    | 570          | 10    | AX097621 Sequence  |
| 7          | 151.6 | 12.6    | 4039         | 85    | AB030001 Homo sapi |
| 8          | 150.6 | 12.5    | 155929       | 69    | AC025574 Homo sapi |

```
9 123.2 10.2 2300 7 AB030002 Sus scrofa
c 10 123.2 10.2 4649 7 AB004061 AB004061 domestic
c 11 62 5.2 7218 10 I66494 Sequence 14
c 12 43.8 3.6 90558 14 NCB24M22 AL390354 Neurospor
c 13 43.8 3.6 258622 66 AC020877 Mus muscu
c 14 43.4 3.6 62336 77 AC087468 AC087468 Homo sapi
c 15 41.8 3.5 1111 53 CNS0763N AL430905 Clone XBA
c 16 41 3.4 209826 73 AC068497 AC068497 Mus muscu
c 17 40.6 3.4 151924 88 AC068722 AC068722 Homo sapi
c 18 40.6 3.4 160482 72 AC055862 AC055862 Homo sapi
c 19 40.6 3.4 186538 79 AL162735 AL162735 Homo sapi
c 20 39.4 3.3 7218 10 I66494 Sequence 14
c 21 39.4 3.3 158802 63 AC015589 AC015589 Homo sapi
c 22 39.2 3.3 193260 69 AC025598 AC025598 Homo sapi
c 23 39 3.2 19020 6 CELR13H8 AF039717 Caenorhab
c 24 39 3.2 109140 83 AP001969 AP001969 Homo sapi
c 25 39 3.2 156628 71 AC027810 AC027810 Homo sapi
c 26 39 3.2 208515 68 AC024619 AC024619 Homo sapi
c 27 38.6 3.2 141268 75 AC074365 AC074365 Homo sapi
c 28 38.6 3.2 158475 83 CNS01DXB AL139194 Homo sapi
c 29 38.4 3.2 1320 95 RNAPN1 AF039889 Rattus no
c 30 38.4 3.2 45408 77 AC087406 AC087406 Homo sapi
c 31 38.4 3.2 182382 76 AC079682 AC079682 Mus muscu
c 32 38 3.2 103270 13 ATFIN20 AL022140 Arabidops
c 33 38 3.2 196247 13 ATCHRIV56 AL161556 Arabidops
c 34 37.8 3.1 132605 68 AC023450 AC023450 Homo sapi
c 35 37.8 3.1 210350 76 AC079536 AC079536 Mus muscu
c 36 37.6 3.1 158749 76 AC079614 AC079614 Homo sapi
c 37 37.6 3.1 202058 64 AC016913 AC016913 Homo sapi
c 38 37.4 3.1 148691 92 HS52D1 Z96811 Human DNA s
c 39 37.4 3.1 162213 71 AC027674 AC027674 Homo sapi
c 40 37.4 3.1 165888 72 AC053506 AC053506 Homo sapi
c 41 37.4 3.1 189631 61 AC009920 AC009920 Homo sapi
c 42 37.4 3.1 208342 66 AC020907 AC020907 Homo sapi
c 43 37.2 3.1 97375 90 AL353898 AL353898 Human DNA
c 44 37.2 3.1 110000 60 AC008749_2 Continuation (3 of
c 45 37.2 3.1 154442 79 AL161915 Homo sapi
```

ALIGNMENTS

```
RESULT 1
AX097623 LOCUS AX097623 1203 bp DNA PAT 30-MAR-2001
DEFINITION Sequence 3 from Patent WO0118051.
ACCESSION AX097623
VERSION AX097623.1 GI:13514264
KEYWORDS unclassified.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1203)
AUTHORS Oppmann,B., de waal Malefyt,R., Rennick,D.M., Kastelein,R.A.,
Wiekowski,M.T., Lira,S.A. and Narula,S.K.
TITLE Mammalian interleukin-12 p40 and interleukin b30. Combinations
thereof. Antibodies. Uses in pharmaceutical compositions
JOURNAL Patent: WO 0118051-A 3 15-MAR-2001;
SCHERING CORPORATION (US)
FEATURES Location/Qualifiers
source 1. 1203
/organism="unclassified"
/db_xref="taxon:32644"
/notes="surmised Mus sp."
113..703
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC35430.1"
/db_xref="GI:13514265"
/translators="MGLDRAVINLWLLPWVTOGLAVPRSSPDWAOOQLSRNLCMLA
WNAHAPAGHNLREDEETKNVPRIQCEDCDPQGLKDNQFCLQRIRQGLATYK
HLRSDIFKGEPLLDPSPMEQHTSLGLSLQQLPDPHPRETQPMPSLSSQQQWRP
LLRSKILSLQAFALAAARVFAHGAATLLEPLVPTA"
```

```
mat_peptide 176..700
BASE COUNT 336 a 310 c 312 g 245 t
ORIGIN
Query Match 100.0%; Score 1203; DB 10; Length 1203;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CGCTTAGAAGTCGACATACAGAGTTAGACTCAGAACCAAGAGAGTGGATAGGGGGTCCA 60
Db 1 CGCTTAGAAGTCGACATACAGAGTTAGACTCAGAACCAAGAGAGTGGATAGGGGGTCCA 60
Qy 61 CAGGCGCTGGTGCAGATCACAGAGCCAGCCAGATCTGAGAAGCAGGGAACAAGATGCTGGA 120
Db 61 CAGGCGCTGGTGCAGATCACAGAGCCAGCCAGATCTGAGAAGCAGGGAACAAGATGCTGGA 120
Qy 121 TTGCAGAGCAGTAATAATGCTATGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 180
Db 121 TTGCAGAGCAGTAATAATGCTATGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 180
Qy 181 TAGGAGTAGCAGTCTGACTGGGCTCAGTGGCCAGCAGCTCTCTCGGAATCTCTCATGCT 240
Db 181 TAGGAGTAGCAGTCTGACTGGGCTCAGTGGCCAGCAGCTCTCTCGGAATCTCTCATGCT 240
Qy 241 AGCCTGGAACGCACATGCACCAGCGGGACATATGAATCTTACTAAGAGAAGAAGAGGATGA 300
Db 241 AGCCTGGAACGCACATGCACCAGCGGGACATATGAATCTTACTAAGAGAAGAAGAGGATGA 300
Qy 301 AGAGACTAAAATAATGTGCCCGTATCCAGTGTGAAGATGGTTGTGACCCACAAAGACT 360
Db 301 AGAGACTAAAATAATGTGCCCGTATCCAGTGTGAAGATGGTTGTGACCCACAAAGACT 360
Qy 361 CAAGGACAACAGCAGATTCGCTTGCAGAAAGATCCGCCAAGGCTCGCTTTTATAAGCA 420
Db 361 CAAGGACAACAGCAGATTCGCTTGCAGAAAGATCCGCCAAGGCTCGCTTTTATAAGCA 420
Qy 421 CCTGCTTGACTCTGACATCTTCAAAGGGAGCGCTCTCTACTCCCTGATAGCCCATGGA 480
Db 421 CCTGCTTGACTCTGACATCTTCAAAGGGAGCGCTCTCTACTCCCTGATAGCCCATGGA 480
Qy 481 GCAACTTCACACCTCCCTACTAGGACTCAGCCAACTCTCCAGCCAGAGGATCACCCCG 540
Db 481 GCAACTTCACACCTCCCTACTAGGACTCAGCCAACTCTCCAGCCAGAGGATCACCCCG 540
Qy 541 GGAGACCAACAGATGCCAGCGCTAGTTCAGTCAGCAGTGGCAGCGCCCTTCTCCG 600
Db 541 GGAGACCAACAGATGCCAGCGCTAGTTCAGTCAGCAGTGGCAGCGCCCTTCTCCG 600
Qy 601 TTCCAAGATCCTTCGAAGCCTCCAGGCTTTTGGCCATAGCTCCCGGGTCTTTGGCCA 660
Db 601 TTCCAAGATCCTTCGAAGCCTCCAGGCTTTTGGCCATAGCTCCCGGGTCTTTGGCCA 660
Qy 661 CGGAGCAGCAACTCTGACTGAGCCCTTAGTGCCCAACAGCTTAAGGATGCCAGGTTCCCA 720
Db 661 CGGAGCAGCAACTCTGACTGAGCCCTTAGTGCCCAACAGCTTAAGGATGCCAGGTTCCCA 720
Qy 721 TGGCTACCATGATAAGACTAATCTATCAGCCACACATCTACCACTTAATTAACCCATTA 780
Db 721 TGGCTACCATGATAAGACTAATCTATCAGCCACACATCTACCACTTAATTAACCCATTA 780
Qy 781 GGACTTGTGCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTT 840
Db 781 GGACTTGTGCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTT 840
Qy 841 GAGAAAAAACAACCCAGAGCAGGCTGCTGCTGAGAAAGAGCTGAGAAAGAGA 900
Db 841 GAGAAAAAACAACCCAGAGCAGGCTGCTGCTGAGAAAGAGCTGAGAAAGAGA 900
Qy 901 ATAAGTCTCGAGCCCTTGGCTTGGAGGGGCAACAGCTGCTGCTGAGGGGAAG 960
Db 901 ATAAGTCTCGAGCCCTTGGCTTGGAGGGGCAACAGCTGCTGCTGAGGGGAAG 960
```

```

QY 961 GGGGGGTTGGCATCGAGAACTGTGAGAAAACCCAGAGCATCAGAAAAAGTGGAGCCAGG 1020
Db 961 GGGGGGTTGGCATCGAGAACTGTGAGAAAACCCAGAGCATCAGAAAAAGTGGAGCCAGG 1020
QY 1021 CTTTGGCCATTATCTGTAGAAAAACAAGAAAGGGGGAACATTATCTCTTCTGGGTGGC 1080
Db 1021 CTTTGGCCATTATCTGTAGAAAAACAAGAAAGGGGGAACATTATCTCTTCTGGGTGGC 1080
QY 1081 TCAGGGAATGTCAGATGCCAGTACTCCAGACAGCAGCTGTGTACCTGGCTGTCTGT 1140
Db 1081 TCAGGGAATGTCAGATGCCAGTACTCCAGACAGCAGCTGTGTACCTGGCTGTCTGT 1140
QY 1141 CCTCAGTTCTAACAGATCTAGTCACTAAGAACTAACAGGAGTACCAGAACTGAC 1200
Db 1141 CCTCAGTTCTAACAGATCTAGTCACTAAGAACTAACAGGAGTACCAGAACTGAC 1200
QY 1201 AAA 1203
Db 1201 AAA 1203

RESULT 2
AF301619 1358 bp mRNA ROD 04-DEC-2000
LOCUS Mus musculus interleukin 23 p19 subunit mRNA, complete cds.
ACCESSION AF301619
VERSION AF301619.1 GI:11528340
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1358)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1358)
AUTHORS Oppmann,B., Lesley,R., Blom,B., Timans,J.C., Xu,Y., Hunte,B.,
Vega,F., Yu,N., Wang,J., Singh,K., Zonin,F., Vaisberg,E.,
Churakova,T., Liu,M., Gorman,D., Wagner,J., Zurawski,S., Liu,Y.,
Abrams,J.S., Moore,K.W., Rennick,D., de Waal-Malefyt,R., Hannum,C.,
Bazan,J.F. and Kastelein,R.A.
TITLE Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with
biological activities similar as well as distinct from IL-12
JOURNAL Immunity 13 (5), 715-725 (2000)
MEDLINE 20567322
REFERENCE 2 (bases 1 to 1358)
AUTHORS Kastelein,R.A., Gorman,D., Timans,J.C., Oppmann,B. and Bazan,J.F.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-2000) Molecular Biology, DNAX Research Institute,
901 California Avenue, Palo Alto, CA 94304, USA
FEATURES
source
1..1358
/organism="Mus musculus"
/db_xref="taxon:10090"
CDS
113..703
/note="four helical bundle cytokine; combines with the p40
subunit of IL-12 to form a novel, biologically active,
composite cytokine, IL-23"
/codon_start=1
/product="interleukin 23 p19 subunit"
/protein_id="RAG37231.1"
/db_xref="gi:11528341"
WNAHAPACHMNLREDEETKNVPRICQDGDPOGLKDNQSFQCLRIQGLAFYK
HLLSDIFKGPALPDSPMEQLHTSLGLSOLLQPEDHPRETQMPSLSSSQWQRP
LLRSKILSLQALIAIAARVFAHAATLTELPLVTA"
BASE COUNT 383 a 326 c 338 g 311 t
ORIGIN
Query Match 100.0%; Score 1203; DB 94; Length 1358;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCTTAGAGTCGGACTCAGAGTTAGACTCAGAACCAAGAGGTGGATAGGGGTCCA 60
Db 1 CGCTTAGAGTCGGACTCAGAGTTAGACTCAGAACCAAGAGGTGGATAGGGGTCCA 60
QY 61 CAGGCTTGGTCAGATCAGAGCCAGCCAGATCTGAGAAAGCAGGGAACAAGATGCTGA 120
Db 61 CAGGCTTGGTCAGATCAGAGCCAGCCAGATCTGAGAAAGCAGGGAACAAGATGCTGA 120
QY 121 TTGTCAGAGCAGTAATAATGCTATGCTGTGCTTGGCTGGGTCACTCAGGGCTGGCTGTG 180
Db 121 TTGTCAGAGCAGTAATAATGCTATGCTGTGCTTGGCTGGGTCACTCAGGGCTGGCTGTG 180
QY 181 TAGGAGTAGCAGTCTCTGACTGGGCTCAGTGCCAGCAGCTCTCTCGGAAATCTCTGCA 240
Db 181 TAGGAGTAGCAGTCTCTGACTGGGCTCAGTGCCAGCAGCTCTCTCGGAAATCTCTGCA 240
QY 241 AGCCTGGAACGACATGCACACAGCGGGGACATATGATCTACTAAGAGAAGAGAGATGA 300
Db 241 AGCCTGGAACGACATGCACACAGCGGGGACATATGATCTACTAAGAGAAGAGAGATGA 300
QY 301 AGAGACTTAAAAATAATGTGCCCGCTATCCAGTGTGAAGATGGTTGTGACCCACAAGGACT 360
Db 301 AGAGACTTAAAAATAATGTGCCCGCTATCCAGTGTGAAGATGGTTGTGACCCACAAGGACT 360
QY 361 CAAGGACAAACAGCAGTCTCTGCTTGCAAAAGGATCCGCCAAGCTTGGCTTTTATAAGCA 420
Db 361 CAAGGACAAACAGCAGTCTCTGCTTGCAAAAGGATCCGCCAAGCTTGGCTTTTATAAGCA 420
QY 421 CCTGCTTGACTCTGACATCTTCAAAGGGGAGGCTGCTCTACTCCCTGATAGCCCCATGA 480
Db 421 CCTGCTTGACTCTGACATCTTCAAAGGGGAGGCTGCTCTACTCCCTGATAGCCCCATGA 480
QY 481 GCAACTTTCACACTCCCTACTAGGACTCAGCCAACTCCCTCCAGCCAGGAGATCACCCCG 540
Db 481 GCAACTTTCACACTCCCTACTAGGACTCAGCCAACTCCCTCCAGCCAGGAGATCACCCCG 540
QY 541 GGAGACCCCAACAGATGCCAGCCTGAGTTCTAGTCAGCAGTGGCAGCGCCCTTCTCCG 600
Db 541 GGAGACCCCAACAGATGCCAGCCTGAGTTCTAGTCAGCAGTGGCAGCGCCCTTCTCCG 600
QY 601 TTCCAGAGTCCCTCGAAGCCTCCAGGCTTTTGGCCATAGCTGCCCGGTCTTTGGCCCA 660
Db 601 TTCCAGAGTCCCTCGAAGCCTCCAGGCTTTTGGCCATAGCTGCCCGGTCTTTGGCCCA 660
QY 661 CGGACGACGAACCTCTGACTGAGCCCTTAGTGCCAAAGCAGCTTAAGGATGCCAGGTTCCTCA 720
Db 661 CGGACGACGAACCTCTGACTGAGCCCTTAGTGCCAAAGCAGCTTAAGGATGCCAGGTTCCTCA 720
QY 721 TGGCTACCATGATAGACTTAATCTATCAGCCAGACATCTACCAGTTAATTAACCCATTA 780
Db 721 TGGCTACCATGATAGACTTAATCTATCAGCCAGACATCTACCAGTTAATTAACCCATTA 780
QY 781 GGACTTGTGCTCTTCTTGTGTTTGGTGAAGGGCAAGGACACCATTTATATAA 840
Db 781 GGACTTGTGCTCTTCTTGTGTTTGGTGAAGGGCAAGGACACCATTTATATAA 840
QY 841 GAGAAAAGAAACAAACCCAGAGCAGCTGGCTGAGAAAGAGAGCTGGAGAAAGA 900
Db 841 GAGAAAAGAAACAAACCCAGAGCAGCTGGCTGAGAAAGAGAGCTGGAGAAAGA 900
QY 901 ATAAAGTCTCGAGCCCTTGGCCCTTGGAAAGCGGCAAGCAGCTGCGGTGGCTGAGGGGAAG 960
Db 901 ATAAAGTCTCGAGCCCTTGGCCCTTGGAAAGCGGCAAGCAGCTGCGGTGGCTGAGGGGAAG 960
QY 961 GGGGGGTGGCATCGAGAACTGTGAGAAAACCCAGAGCATCAGAAAAAGTGGAGCCAGG 1020
Db 961 GGGGGGTGGCATCGAGAACTGTGAGAAAACCCAGAGCATCAGAAAAAGTGGAGCCAGG 1020
QY 1021 CTTTGGCCATTATCTGTAGAAAAACAAGAAAGGGGGAACATTATCTCTTCTGGGTGGC 1080
Db 1021 CTTTGGCCATTATCTGTAGAAAAACAAGAAAGGGGGAACATTATCTCTTCTGGGTGGC 1080
QY 1081 TCAGGGAATGTCAGATGCCAGTACTCCAGACAGCAGCTGTGTACCTGGCTGTCTGT 1140
Db 1081 TCAGGGAATGTCAGATGCCAGTACTCCAGACAGCAGCTGTGTACCTGGCTGTCTGT 1140

```

```

Qy 1141 CCCTCAGTCTTAACAGAACTAGTCACTAAGAACTAACAGGACTACCAATACGAACTGAC 1200
|||||
Db 1141 CCCTCAGTCTTAACAGAACTAGTCACTAAGAACTAACAGGACTACCAATACGAACTGAC 1200
Qy 1201 AAA 1203
|||
Db 1201 AAA 1203

RESULT 3
AF301620 1025 bp mRNA PRI 04-DEC-2000
LOCUS Homo sapiens Interleukin 23 p19 subunit mRNA, complete cds.
DEFINITION AF301620
ACCESSION AF301620
VERSION AF301620.1 GI:11528342
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1025)
AUTHORS Oppmann,B., Lesley,R., Blom,B., Timans,J.C., Xu,Y., Hunte,B.,
Vega,F., Yu,N., Wang,J., Singh,K., Zonin,F., Vaisberg,E.,
Churakova,T., Liu,M., Gorman,D., Wagner,J., Zurawski,S., Liu,Y.,
Abrams,J.S., Moore,K.W., Rennick,D., de Waal-Malefyt,R., Hannum,C.,
Bazan,J.F. and Kastelein,R.A.
TITLE Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with
biological activities similar as well as distinct from IL-12
JOURNAL Immunity 13 (5), 715-725 (2000)
MEDLINE 20567322
REFERENCE 2 (bases 1 to 1025)
AUTHORS Kastelein,R.A., Gorman,D., Timans,J.C., Oppmann,B. and Bazan,J.F.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-2000) Molecular Biology, DNAX Research Institute,
901 California Avenue, Palo Alto, CA 94304, USA
FEATURES
source
1. .1025
/organism="Homo sapiens"
/db_xref="taxon:9606"
151..720
CDS
/note="combines with the p40 subunit of IL-12 to form a
novel, biologically active, composite cytokine, IL-23"
/codon_start=1
/product="interleukin 23 p19 subunit"
/protein_id="AAG37232.1"
/db_xref="GI:11528343"
/translation="MLGSRAVMLLLPWTAGRAVPGSSPAWTCQQLSKLCTLA
WSAHLVGHMDLREEDTNDVPHIQGDGCDPQGLRDNSTQFLRIHQGLIFYEK
LLGSDIFTEPSPVLAQSLHSLQLPQEGHHWETQIPLSPSPQWRLL
LRFKILRSIQAFVAARVFAHGAATLSP"
BASE COUNT 262 a 252 c 274 g 237 t
ORIGIN

Query Match 36.7%; Score 441.4; DB 89; Length 1025;
Best Local Similarity 77.5%; Pred. No. 4.5e-124;
Matches 601; Conservative 0; Mismatches 161; Indels 13; Gaps 5;

Qy 18 ACAGAGTACAGTACAGAACCAAGGAGGTGATAGGGGGTCCA-CAGGCCTGGTGCAGAT 76
|||||
Db 55 AGAAGTACAGCTCAAGCAAGTGAAGTGGCAGAGATTCCACGAGACTGTGTCAAGG 114
|||||
Qy 77 CACAGAGCCAGCCAGTCTGAGAGCAGGGAACAAGATGCTGGATTGCAGAGCAGTAATA 136
|||||
Db 115 CGCAGAGCCAGCAGATTGAGAGAGAGGCAAAAGATGCTGGGACAGAGCTGTAATG 174
Qy 137 ATGCTATGGCTGTGGCTGGCTCACTCAGGGCTGGCTGTGGCTTAGAGTAGCAGTCTCT 196
|||||
Db 175 CTGCTGTGTGCTGCTGCTGGCAGAGCTCAGGGCAGAGCTGTGGCTGGGGGCGAGCAGCCT 234
Qy 197 GACTGGGCTCAGTCCAGCAGCTCTCTCGGAATCTCTGCATGCTAGCTGGAACGCACAT 256
|||||

```

## RESULT 4

```

AB030000 1026 bp mRNA PRI 04-APR-2000
LOCUS Homo sapiens mRNA for SGRF, complete cds.
DEFINITION AB030000
ACCESSION AB030000
VERSION AB030000.1 GI:7416072
KEYWORDS SGRF.
SOURCE Homo sapiens fetal spleen cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1026)
AUTHORS Hirata,Y. and Kosuge,Y.
TITLE SGRF: a novel member of the IL-6/G-CSF family
JOURNAL Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 1026)
AUTHORS Hirata,Y. and Kosuge,Y.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-1999) to the DBJ/EMBL/GenBank databases. Yuichi
Hirata, Chugai Research Institute for Molecular Medicine, Gene
Search Program, 153-2 Nagai, Nihari-mura, Ibaraki 300-4101, Japan
(E-mail:hiratayu@climmed.com, Tel:81-298-30-6211(ex.288),
Fax:81-298-30-6270)
FEATURES
source
1. .1026
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="fetal"
/tissue_type="spleen"
144..713
/gene="SGRF"
144..713
/gene="SGRF"
/codon_start=1

```











RESULT 12  
 NCB24M22/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 JOURNAL  
 TITLE  
 JOURNAL  
 COMMENT  
 FEATURES  
 source  
 misc\_feature  
 repeat\_region  
 gene  
 CDS  
 exon  
 intron



QY 880 GAAAGGAGCTGGAGAGAAGATAAAG 906  
| | | | | | | | | | | | | | | | | | |  
Db 84150 TGGAGAGCTGGAGAGAAGGACAAG 84124

RESULT 13  
AC020877/c  
LOCUS  
DEFINITION Mus musculus clone RP23-344L20, WORKING DRAFT SEQUENCE, 37  
ACCESSION AC020877  
VERSION AC020877.2 GI:7340303  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
JOURNAL 1 (bases 1 to 258622)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 258622)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Mar 29, 2000 this sequence version replaced gi:5686431.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----Summary Statistics  
Consensus quality: 140189 bases at least Q40  
Consensus quality: 191513 bases at least Q30  
Consensus quality: 208421 bases at least Q20  
Estimated insert size: 258622; sum-of-contigs estimation  
Estimated insert size: 216000; pulse field gel estimation  
Quality coverage: 4.19x in Q20 bases; pulse field gel estimation  
Quality coverage: 3.50x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 97 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 1069: contig of 1069 bp in length  
\* gap of unknown length  
\* 1070 2085: contig of 1016 bp in length  
\* gap of unknown length  
\* 2086 3153: gap of unknown length  
\* contig of 1068 bp in length  
\* 3154 4722: gap of unknown length  
\* contig of 1569 bp in length  
\* 4723 5728: gap of unknown length  
\* contig of 1006 bp in length  
\* 5729 6764: gap of unknown length  
\* contig of 1036 bp in length  
\* 6765 7830: gap of unknown length  
\* contig of 1066 bp in length  
\* 7831 8859: gap of unknown length  
\* contig of 1029 bp in length  
\* 8860 9865: gap of unknown length  
\* contig of 1006 bp in length  
\* 9866 10869: gap of unknown length  
\* contig of 1004 bp in length  
\* 10870 11962: gap of unknown length  
\* contig of 1093 bp in length  
\* 11963 13013: gap of unknown length  
\* contig of 1051 bp in length  
\* gap of unknown length  
\* 13014 14338: contig of 1325 bp in length  
\* gap of unknown length  
\*  
\* 14339 15557: gap of unknown length  
\* contig of 1219 bp in length  
\* 15558 16667: gap of unknown length  
\* contig of 1110 bp in length  
\* 16668 17748: gap of unknown length  
\* contig of 1081 bp in length  
\* 17749 19089: gap of unknown length  
\* contig of 1341 bp in length  
\* 19090 20675: gap of unknown length  
\* contig of 1586 bp in length  
\* 20676 22436: gap of unknown length  
\* contig of 1761 bp in length  
\* 22437 23683: gap of unknown length  
\* contig of 1247 bp in length  
\* 23684 25215: gap of unknown length  
\* contig of 1532 bp in length  
\* 25216 26431: gap of unknown length  
\* contig of 1216 bp in length  
\* 26432 27491: gap of unknown length  
\* contig of 1060 bp in length  
\* 27492 28737: gap of unknown length  
\* contig of 1246 bp in length  
\* 28738 29742: gap of unknown length  
\* contig of 1005 bp in length  
\* 29743 30947: gap of unknown length  
\* contig of 1205 bp in length  
\* 30948 31992: gap of unknown length  
\* contig of 1045 bp in length  
\* 31993 33108: gap of unknown length  
\* contig of 1116 bp in length  
\* 33109 34435: gap of unknown length  
\* contig of 1327 bp in length  
\* 34436 35651: gap of unknown length  
\* contig of 1216 bp in length  
\* 35652 37355: gap of unknown length  
\* contig of 1704 bp in length  
\* 37356 38598: gap of unknown length  
\* contig of 1243 bp in length  
\* 38599 39644: gap of unknown length  
\* contig of 1046 bp in length  
\* 39645 41275: gap of unknown length  
\* contig of 1631 bp in length  
\* 41276 42381: gap of unknown length  
\* contig of 1106 bp in length  
\* 42382 43462: gap of unknown length  
\* contig of 1081 bp in length  
\* 43463 44822: gap of unknown length  
\* contig of 1360 bp in length  
\* 44823 45961: gap of unknown length  
\* contig of 1139 bp in length  
\* 45962 47534: gap of unknown length  
\* contig of 1573 bp in length  
\* 47535 49241: gap of unknown length  
\* contig of 1707 bp in length  
\* 49242 50343: gap of unknown length  
\* contig of 1102 bp in length  
\* 50344 51833: gap of unknown length  
\* contig of 1490 bp in length  
\* 51834 52901: gap of unknown length  
\* contig of 1068 bp in length  
\* 52902 54595: gap of unknown length  
\* contig of 1694 bp in length  
\* 54596 56094: gap of unknown length  
\* contig of 1499 bp in length  
\* 56095 57171: gap of unknown length  
\* contig of 1077 bp in length  
\* 57172 58192: gap of unknown length  
\* contig of 1021 bp in length  
\* 58193 59434: gap of unknown length  
\* contig of 1242 bp in length  
\* 59435 60648: gap of unknown length  
\* contig of 1214 bp in length  
\* gap of unknown length



\* 1527 1626: gap of 100 bp  
\* 1627 2364: contig of 738 bp in length  
\* 2365 2464: gap of 100 bp  
\* 2465 3198: contig of 734 bp in length  
\* 3199 3298: gap of 100 bp  
\* 3299 3959: contig of 661 bp in length  
\* 3960 4059: gap of 100 bp  
\* 4060 4778: contig of 719 bp in length  
\* 4779 4878: gap of 100 bp  
\* 4879 5619: contig of 741 bp in length  
\* 5620 5719: gap of 100 bp  
\* 5720 6435: contig of 716 bp in length  
\* 6436 6535: gap of 100 bp  
\* 6536 7261: contig of 726 bp in length  
\* 7262 7361: gap of 100 bp  
\* 7362 8081: contig of 720 bp in length  
\* 8082 8181: gap of 100 bp  
\* 8182 8914: contig of 733 bp in length  
\* 8915 9014: gap of 100 bp  
\* 9015 9744: contig of 730 bp in length  
\* 9745 9844: gap of 100 bp  
\* 9845 10570: contig of 726 bp in length  
\* 10571 10670: gap of 100 bp  
\* 10671 11396: contig of 726 bp in length  
\* 11397 11496: gap of 100 bp  
\* 11497 12222: contig of 726 bp in length  
\* 12223 12322: gap of 100 bp  
\* 12323 13050: contig of 728 bp in length  
\* 13051 13150: gap of 100 bp  
\* 13151 13865: contig of 715 bp in length  
\* 13866 13965: gap of 100 bp  
\* 13966 14675: contig of 710 bp in length  
\* 14676 14775: gap of 100 bp  
\* 14776 15517: contig of 742 bp in length  
\* 15518 15617: gap of 100 bp  
\* 15618 16330: contig of 713 bp in length  
\* 16331 16430: gap of 100 bp  
\* 16431 17142: contig of 712 bp in length  
\* 17143 17242: gap of 100 bp  
\* 17243 17967: contig of 725 bp in length  
\* 17968 18067: gap of 100 bp  
\* 18068 18783: contig of 716 bp in length  
\* 18784 18883: gap of 100 bp  
\* 18884 19604: contig of 721 bp in length  
\* 19605 19704: gap of 100 bp  
\* 19705 20428: contig of 724 bp in length  
\* 20429 20528: gap of 100 bp  
\* 20529 21215: contig of 687 bp in length  
\* 21216 21315: gap of 100 bp  
\* 21316 22044: contig of 729 bp in length  
\* 22045 22144: gap of 100 bp  
\* 22145 22875: contig of 731 bp in length  
\* 22876 22975: gap of 100 bp  
\* 22976 23696: contig of 721 bp in length  
\* 23697 23796: gap of 100 bp  
\* 23797 24523: contig of 727 bp in length  
\* 24524 24623: gap of 100 bp  
\* 24624 25337: contig of 714 bp in length  
\* 25338 25437: gap of 100 bp  
\* 25438 26169: contig of 732 bp in length  
\* 26170 26269: gap of 100 bp  
\* 26270 26978: contig of 709 bp in length  
\* 26979 27078: gap of 100 bp  
\* 27079 27772: contig of 694 bp in length  
\* 27773 27872: gap of 100 bp  
\* 27873 28599: contig of 727 bp in length  
\* 28600 28699: gap of 100 bp  
\* 28700 29405: contig of 706 bp in length  
\* 29406 29505: gap of 100 bp  
\* 29506 30231: contig of 726 bp in length  
\* 30232 30331: gap of 100 bp  
\* 30332 31052: contig of 721 bp in length  
\* 31053 31152: gap of 100 bp

\* 31153 31884: contig of 732 bp in length  
\* 31885 31984: gap of 100 bp  
\* 31985 32698: contig of 714 bp in length  
\* 32699 32798: gap of 100 bp  
\* 32799 33519: contig of 721 bp in length  
\* 33520 33619: gap of 100 bp  
\* 33620 34367: contig of 748 bp in length  
\* 34368 34467: gap of 100 bp  
\* 34468 35194: contig of 727 bp in length  
\* 35195 35294: gap of 100 bp  
\* 35295 36021: contig of 727 bp in length  
\* 36022 36121: gap of 100 bp  
\* 36122 36834: contig of 713 bp in length  
\* 36835 36934: gap of 100 bp  
\* 36935 37649: contig of 715 bp in length  
\* 37650 37749: gap of 100 bp  
\* 37750 38466: contig of 717 bp in length  
\* 38467 38566: gap of 100 bp  
\* 38567 39289: contig of 723 bp in length  
\* 39290 39389: gap of 100 bp  
\* 39390 40121: contig of 732 bp in length  
\* 40122 40221: gap of 100 bp  
\* 40222 40950: contig of 729 bp in length  
\* 40951 41050: gap of 100 bp  
\* 41051 41777: contig of 727 bp in length  
\* 41778 41877: gap of 100 bp  
\* 41878 42617: contig of 740 bp in length  
\* 42618 42717: gap of 100 bp  
\* 42718 43431: contig of 714 bp in length  
\* 43432 43531: gap of 100 bp  
\* 43532 44252: contig of 721 bp in length  
\* 44253 44352: gap of 100 bp  
\* 44353 45064: contig of 712 bp in length  
\* 45065 45164: gap of 100 bp  
\* 45165 45872: contig of 708 bp in length  
\* 45873 45972: gap of 100 bp  
\* 45973 46667: contig of 695 bp in length  
\* 46668 46767: gap of 100 bp  
\* 46768 47497: contig of 730 bp in length  
\* 47498 47597: gap of 100 bp  
\* 47598 48326: contig of 729 bp in length  
\* 48327 48426: gap of 100 bp  
\* 48427 49164: contig of 738 bp in length  
\* 49165 49264: gap of 100 bp  
\* 49265 49988: contig of 724 bp in length  
\* 49989 50088: gap of 100 bp  
\* 50089 50813: contig of 725 bp in length  
\* 50814 50913: gap of 100 bp  
\* 50914 51645: contig of 732 bp in length  
\* 51646 51745: gap of 100 bp  
\* 51746 52462: contig of 717 bp in length  
\* 52463 52562: gap of 100 bp  
\* 52563 53280: contig of 718 bp in length  
\* 53281 53380: gap of 100 bp  
\* 53381 54092: contig of 712 bp in length  
\* 54093 54192: gap of 100 bp  
\* 54193 54908: contig of 716 bp in length  
\* 54909 55008: gap of 100 bp  
\* 55009 55733: contig of 725 bp in length  
\* 55734 55833: gap of 100 bp  
\* 55834 56569: contig of 736 bp in length  
\* 56570 56669: gap of 100 bp

Query Match 3.6%; Score 43.4; DB 77; Length 62336;  
Best Local Similarity 34.9%; Pred. No. 0.12;  
Matches 149; Conservative 0; Mismatches 278; Indels 0; Gaps 0;

QY 633 TGGCCATAGCTGCCCCGGGTCTTTGGCCACGAGGACCACTCTGACTGAGCCCTTAGTCG 692

DB 59214 TTGCTTAAAGGGAAGCGGTGTTTTTCNACCCCAATGGAAGGGTTANATACCAAGTTGC 59155

QY 693 CAACAGCTTAAGGATGCCCCAGGTTCCCATGGCTFACCATGATAGATAAGATAATCTATCAGCCC 752









| Result No. | Score | Query |      | Length | DB     | ID | Description         |
|------------|-------|-------|------|--------|--------|----|---------------------|
|            |       | Match | %    |        |        |    |                     |
| 1          | 1203  | 100.0 | 1203 | 20     | Z08866 |    | Mouse interleukin   |
| 2          | 1203  | 100.0 | 1203 | 20     | X17787 |    | Mouse interleukin   |
| 3          | 1203  | 100.0 | 1203 | 21     | A52578 |    | Murine interleukin  |
| 4          | 439.8 | 36.6  | 1026 | 21     | Z37262 |    | SGRF coding sequen  |
| 5          | 439.8 | 36.6  | 1055 | 22     | C84306 |    | Human EXCS' encodin |
| 6          | 439.8 | 36.6  | 1067 | 21     | A16886 |    | Human secreted pro  |
| 7          | 391   | 32.5  | 570  | 20     | Z08865 |    | Human interleukin   |
| 8          | 391   | 32.5  | 570  | 20     | X17786 |    | Human interleukin   |
| 9          | 391   | 32.5  | 570  | 21     | A52577 |    | Human interleukin   |
| 10         | 203.2 | 16.9  | 412  | 21     | A42676 |    | Human secreted exp  |
| 11         | 151.6 | 12.6  | 2398 | 21     | Z37263 |    | SGRF coding sequen  |

XX (SCHE ) SCHERING CORP.  
XX PA Kastelein RA, Mattson JD, McClanahan TK;  
XX PI WPI; 1999-527306/44.  
XX DR P-PSDB; Y29784.  
XX PT New receptor subunits useful in the treatment inflammatory disorders  
XX PS Claim 16; Page 27-28; 133pp; English.  
XX CC The present invention describes a composition (I) comprising DNAX  
CC cytokine receptor subunit I (Dcrs1) protein and DNAX soluble receptor  
CC subunit I (Dcrs1) protein, which together encode a new mammalian  
CC cytokine-related receptor (R), or Dcrs1 and interleukin B30 (IL-B30)  
CC proteins, or Dcrs1 and IL-B30 proteins. (I) comprising Dcrs1 and Dcrs1  
CC is useful for screening for ligands (i.e. agonists/antagonists) from  
CC a library of compounds, which are useful for modulating the physiology  
CC or development of a cell or tissue culture e.g. inflammatory responses,  
CC innate immunity and/or morphogenic development. (R), antibodies and  
CC ligands are useful for treatment of conditions, especially immunological  
CC disorders, associated with conditions exhibiting abnormal expression of  
CC (R). (R) is useful as a phosphate labeling enzyme to label substrates,  
CC and the subunits Dcrs1 and Dcrs1 are useful as immunogens for generating  
CC antibodies, or as antigens for binding antibodies. Nucleic acids  
CC encoding (R) are useful for identifying related DNAs and mRNAs, and  
CC variants from other individuals or species. The present sequence  
CC encodes the specifically claimed mouse IL-B30, for use in the  
CC composition of the present invention.  
XX SQ Sequence 1203 BP; 336 A; 310 C; 312 G; 245 T; 0 other;

Query Match 100.0%; Score 1203; DB 20; Length 1203;  
Best Local Similarity 100.0%; Pred. NO. 0;  
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCTTAGAAGTCGGACTACAGAGTTAGACTCAGAACCAAGAGGTGGATAGGGGTCCA 60  
Db 1 cgcttagaagtcggactacagagttagactcagaacaaaggagtgatagggtcca 60  
Qy 61 CAGGCTGGTCAGATCAGACAGCCAGCCAGATCTGAGAACGAGGGAACAGATGCTGGA 120  
Db 61 caggcctggtcagatcacagacagccagcagatctgagaagcagggaacaagatgctgga 120  
Qy 121 TTGAGAGCAGTAATATGCTATGCTGTTCCTGGTGGTCACTCAGGCTGGCTGTGCC 180  
Db 121 ttgagagcagtaataatgctatgctgttgcctgggtggtcaactcaggcctggctgtgcc 180  
Qy 181 TAGGAGTAGCAGTCTGCTGCTGGCTCAGTCAGCAGCTCTCTCGGAATCTCTGCATGCT 240  
Db 181 taggagtagcagtcctgactggtggtcagtcagtcagcagcctctcggaaatcctcagtcgt 240  
Qy 241 AGCCTGGAACGCATGACACAGCGGGACATATGAATCTACTAAGAGAAGAGGATGA 300  
Db 241 agcctggaacgcatacaccagcggacatatgaatctactaagaagaagagatga 300  
Qy 301 AGAGCTAAAAATAATGTGCCCTATCCAGTGTGAAGATGTTGTGACCCACAGGACT 360  
Db 301 agagactaaaaataatgtgccctatccagtgtagaatggtgtgaccacagact 360  
Qy 361 CAAGGACAACAGCAGTCTGCTGTCGAAAGATCCGCAAGGTCTGGCTTTTATAAGCA 420  
Db 361 caaggacaacagcagttctgcttgaaaggatccgcaaggctcgtcttttataagca 420  
Qy 421 CCTGCTGACTCTGACATCTTCAAGGGAGGCTGCTGCTACTCTCCTGATGCCCATGGA 480  
Db 421 cctgctgactctgacatcttcaaggaggcctgctctactcctctgatagcccatgga 480  
Qy 481 GCACTTTCACACTTCCCTACTAGACTCAGCCAACTCTCCAGCCAGGATCACCCCG 540  
Db 481 gcaacttcaacctccctactaggaactcagccaaactcctccagccagatcaccccg 540

Qy 541 GGAGACCAACAGATGCCAGCCTGAGTTCTAGTCAGCAGTGGCAGGCCCTTCTCCG 600  
Db 541 ggagaccaacagatgccagcctgagttctagtcagcagtgggcagcgccctctccg 600  
Qy 601 TTCCAAGATCCTCGAAGCCTCCAGGCTTTTGGCCATAGCTGCCGGGTCTTTGGCCA 660  
Db 601 ttccaagatcctcgaaagcctccaggccttttggccatagctgccgggtctttggcca 660  
Qy 661 CGGAGCAGCAACTCTGACTGAGCCCTTAGTCCCAACAGCTTAAGGATGCCAGGTTCCA 720  
Db 661 cggagcagcaactctgactgagccttagtgcacacagcttaaggatgccaggttccca 720  
Qy 721 TGGCTACCATGATAAGACTAATCTATCAGCCAGCACATCTACCATTAATTAACCCATTA 780  
Db 721 tggctaccatgataagactaatactatcagccagacatctaccagttaataaacccatta 780  
Qy 781 GGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
Db 781 ggacttgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 840  
Qy 841 GAGAAAGAAACAAACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 900  
Db 841 gagaaagaaacaaacccagcagcagcagcagcagcagcagcagcagcagcagcagcag 900  
Qy 901 ATAAAGTCTCAGGCTTGGCTTGGAGCGGGCAAGCAGCTGGCTGGCTGGCTGGCTGG 960  
Db 901 ataaagtctcagccttggccttggagcggcggaagcagcagcagcagcagcagcagcag 960  
Qy 961 GGGCGGTGGCATCGAAGAACTGTGAGAAACCCAGCAGCATCAGAAAAGTGAGCCGAG 1020  
Db 961 gggcggtggcatcgaaactgtgagaaacccagcagcatcagaaaagtgagccagcag 1020  
Qy 1021 CTTTGGCATTATCTGTAAGAAACAAAGGGAACATTATCTTCTCGGTGGC 1080  
Db 1021 ctttggcattatctgtaagaaacaaaggaaggggaacattatcttctcgtggc 1080  
Qy 1081 TCAGGGAATGTGAGATGCAGTACTCTCAGCAGCAGCAGCTCTGCTGCTGCTGCTGCT 1140  
Db 1081 tcagggaatgtgagatgcagactctcagcagcagcagcagcagcagcagcagcagcagc 1140  
Qy 1141 CCCTCAGTTCTAAGCAATCTAGTACTAGAACTACAGGACTACCAATACGAACTGAC 1200  
Db 1141 ccctcagttctaaagaatctagtcactaagaactaacaggaactaacgaactgac 1200  
Qy 1201 AAA 1203  
Db 1201 aaa 1203

RESULT 2  
X17787  
ID X17787 standard; cDNA; 1203 BP.  
XX X17787;  
AC X17787;  
XX 21-MAY-1999 (first entry)  
XX Mouse interleukin-B30 (IL-B30) polypeptide encoding cDNA.  
XX Cytokine; interleukin-B30; IL-B30; forensic science; cell proliferation;  
XX Inflammatory condition; drug screening; mouse; ss.  
XX Mus sp.  
XX Key Location/Qualifiers  
XX CDS 113..703  
XX /\*tag= a  
XX /\*product= "IL-B30"  
XX sig\_peptide 113..175  
XX /\*tag= b  
XX mat\_peptide 176..700  
XX /\*tag= c



XX FH Key Location/Qualifiers  
 FT CDS 113..703  
 FT /\*tag= a  
 FT /product= "Murine IL-B30 precursor"  
 FT 113..175  
 FT sig\_peptide  
 FT /\*tag= b  
 FT 176..700  
 FT mat\_peptide  
 FT /\*tag= c  
 FT /product= "Mature murine IL-B30"  
 XX  
 PN US060284-A.  
 XX  
 XX 09-MAY-2000.  
 XX  
 XX 24-JUL-1998; 98US-0122443.  
 XX  
 XX 25-JUL-1997; 97US-0053765.  
 XX  
 XX (SCHE ) SCHERING CORP.  
 XX  
 XX Bazan JF;  
 XX  
 XX WPI; 2000-364420/31.  
 DR P-PSDB; B01982.  
 XX  
 XX Novel recombinant DNA encoding cytokines especially interleukin-B30  
 PT useful as probes or primers for diagnosing immune disorders including  
 PT autoimmune or chronic inflammatory conditions -  
 XX  
 PS Claim 3; Column 7-10; 32pp; English.  
 XX  
 CC This sequence represents cDNA encoding murine interleukin-B30 (IL-B30).  
 CC IL-B30 is a novel cytokine, exhibiting significant homology to IL-6 and  
 CC G-CSF (granulocyte colony stimulating factor). Cytokines play a critical  
 CC role in signalling between immune or other cells during an immune  
 CC response. The precise role of IL-B30 is not yet known - it is likely to  
 CC have either a stimulatory or an inhibitory effect on haematopoietic cells  
 CC such as T-cells, B-cells, natural killer (NK) cells and macrophages.  
 CC Alternatively, it may affect vascular physiology or development, or have  
 CC neuronal effects. IL-B30, its fragments, IL-B30 nucleotides, agonists and  
 CC antagonists are useful in the diagnosis and treatment of disorders  
 CC associated with abnormal expression or activity of IL-B30 e.g.,  
 CC inflammatory or autoimmune disorders. Nucleotides encoding IL-B30 are  
 CC useful for recombinant expression of IL-B30 in a host cell, and as a  
 CC source of probes and primers. The IL-B30 probes and primers can be used  
 CC to detect levels of IL-B30 expression in samples from patients suspected  
 CC of having an inflammatory or autoimmune disorder. IL-B30 nucleotides may  
 CC also be used to identify homologous genes in other species. IL-B30  
 CC protein or its fragments are useful as antigens for raising antibodies to  
 CC various linear and conformational epitopes. Such antibodies may be used  
 CC to detect levels of IL-B30 protein in a sample.  
 XX  
 XX Sequence 1203 BP; 336 A; 310 C; 312 G; 245 T; 0 other;  
 SQ

Query Match 100.0%; Score 1203; DB 21; Length 1203;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCTTAGAGTCGACATACAGAGTTAGACTCAGAACCAAGGAGGTGGATAGGGGTCCA 60  
 |||||  
 DB 1 cgcttagaagtcgactacagagtagtagactcagaacaaagggtggataggggtcca 60  
 |||||  
 QY 61 CAGGCCGTGGTCATCACAGAGCCAGCCAGATCTGAGAGCAGGGAACAGAGTCTGGGA 120  
 |||||  
 DB 61 caggccctggctcagatcacagagcagccagatctggaagcagggaacagaatgctcggga 120  
 |||||  
 QY 121 TTGCAGAGCAGTAATAATGCTATGGCTGTGGCTGGCTCAGTCAGGCCCTGGCTGTGCC 180  
 |||||  
 DB 121 ttgcagagcagtaataatgctatggtgtggcctggcctcagtcagggcctggctgtgcc 180  
 |||||  
 QY 181 TAGGAGTAGCAGTCTCTGACGTGGGCTCAGTGGCCAGCAGCTCTCTCGGAATCTCTGCATGCT 240

Db 181 taggagtagcagtcctgactgggctcagtgccagcagctctctcggaatctctgctgct 240  
 |||||  
 QY 241 AGCCTGGAACGCACATGCACAGCGGACATATCAATCTACTAGAGAGAGAGGATGA 300  
 |||||  
 DB 241 agcctggaacgcacatgcaccagcgggacatgaaactactaagaagaagaaggatga 300  
 |||||  
 QY 301 AGAGACTAAAATAATGTGCCCGTATCCAGTGTGAAGATGGTGTGACCCACACAGGACT 360  
 |||||  
 DB 301 agagactaaaataatgtgcccgtatccagtggaagatggtgtgaccacaaggact 360  
 |||||  
 QY 361 CAAGGACAACAGCAGTCTCTGCTTTCGAAAGATCCGCAAGGTCTGCTTTTATAAGCA 420  
 |||||  
 DB 361 caaggacaacagcagttctgtctgcaagatccgcaaggtctggtcttttataagca 420  
 |||||  
 QY 421 CCTGCTTACTCTGACATCTTCAAGGGGAGCCTGCTCTACTCCCTGATAGCCCATGGA 480  
 |||||  
 DB 421 cctgcttactctgacatcttcaaggaggagcctgctactcctctgatagcccatgga 480  
 |||||  
 QY 481 GCAACTTCACACCTCCCTACTAGGACTCAGCAACTCTCTCCAGCCAGAGGATACCCCG 540  
 |||||  
 DB 481 gcaacttcacacctccctactactagactcagccaactcctccagccagaggataccccc 540  
 |||||  
 QY 541 GGAGACCCCAACAGATGCCAGCCTGAGTCTTAGTCAGCAGTGGCGGCCCTTCTCCG 600  
 |||||  
 DB 541 ggagacccaacagatgccagcctgagttctagtcagcagtgggcagcgccccctctccg 600  
 |||||  
 QY 601 TTCCAAGATCTTCGAGAGCCTCCAGGCTTTTGGCCATAGTCGCCCGGGTCTTTGGCCA 660  
 |||||  
 DB 601 ttccaagatccttcgaagcctccaggccttttggccatagctgcccgggtcttggcca 660  
 |||||  
 QY 661 CGGAGCAGCAACTCTGACTGAGCCTTTAGTGCCCAACAGCTTAAGGATGCCAGTTCCCA 720  
 |||||  
 DB 661 cggagcagcaactctgactgagcccttagtgccaacagcttaaggatgccaggttccca 720  
 |||||  
 QY 721 TGGTACCATGATGAAGACTTAATCTATCAGGCCAGACATCTACCAAGTTAATTAACCCATTA 780  
 |||||  
 DB 721 tggtagcatgataagaactaatctatcagccagacatctaccagtttaattaaacccatta 780  
 |||||  
 QY 781 GGACTTCTGCTGTCTTGTCTTCTGTTTTCCTGAGGCAAGGCAAGACACCATTTATAA 840  
 |||||  
 DB 781 ggaacttctgctgtcttctgttcttctgttcttctgaggggcaaggacacattatataa 840  
 |||||  
 QY 841 GAGAAAAAACAACCCAGAGCAGCAGCTGGCTGAGAGAAAGAGCTGAGAGGAAGA 900  
 |||||  
 DB 841 gagaaaaaacaacccccagagcagcgctggtctagagaaaggagctggagaaga 900  
 |||||  
 QY 901 ATAAAGTCTCGAGCCCTTGGCTTGGAGGGGCAAGCAGCTGGCTGGCTGAGGGGAAG 960  
 |||||  
 DB 901 ataaagtctcgagcccttggccttggaggggcaagcagctgctggtgaggggaag 960  
 |||||  
 QY 961 GGGCGGTGGCATCGAAGAACTGTGAGAAAACCCAGAGCATCAGAAAAAGTGAGCCCGCA 1020  
 |||||  
 DB 961 gggcggtggcatcgagaaactgtgagaaaacccagagcagcagaaaaagtggagccagg 1020  
 |||||  
 QY 1021 CTTTGGCCATTATCTGTAAGAAAAACAAGAAAGGGAACATTTACTTTCTCGGTGGC 1080  
 |||||  
 DB 1021 ctttggccattatctgtaagaaaaacaagaaagggaacattatacttctctggtggc 1080  
 |||||  
 QY 1081 TCAGGGAATGTGCAGATGCACAGTACTTCCAGAGCAGAGCTCTGTACTGCTCTGT 1140  
 |||||  
 DB 1081 tcagggaatgtgcagatgcacagtagtactccagacagcagctctgtactcgtcgtctgt 1140  
 |||||  
 QY 1141 CCCTCAGTTCTTAACAGAACTCTAGTCACTAAGAACTAACAGSACTACCAATACGAACCTGAC 1200  
 |||||  
 DB 1141 ccctcagttcttaacagaatctagtcactaagaactaacaggactaccaatacgaactgac 1200  
 |||||  
 QY 1201 AAA 1203  
 DB 1201 aaa 1203

RESULT 4



The invention provides human extracellular signaling molecules (EXCS) and polynucleotides which identify and encode EXCS. EXCS can be expressed by standard recombinant methodology. The amino acid and nucleic acid sequences of EXCS are useful for diagnosing, treating and preventing infections and gastrointestinal (peptic ulcer, dysphagia, pancreatitis), neurological (e.g. epilepsy, ischemic cerebrovascular disease, stroke), reproductive (infertility, ovulatory defects, endometriosis), autoimmune /inflammatory (actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease), and cell proliferative disorders including cancers (of the breast, adrenal gland, bone). They may also be used to treat fatal familial insomnia, nutritional and metabolic diseases of the nervous system, myopathies, mental disorders (anxiety, schizophrenia, mood), as well as infections caused by parasites (malaria, leishmania, trypanosoma), viral (adenovirus, coronavirus, flavivirus), bacterial (e.g. pneumococcus, staphylococcus, bacillus), and fungal (aspergillus, blastomycosis, dermatophytes) agents. The nucleic acids, polypeptides, antagonists, agonists, pharmaceutical compositions, and antibodies may also be used for treating or preventing disorders associated with increased or decreased expression or activity of EXCS. EXCS polynucleotides may also be used to detect and quantify gene expression in biopsied tissues in which expression of EXCS may be correlated with the disease, to determine presence or excess expression of EXCS, to monitor regulation of EXCS levels during therapeutic intervention, to detect the presence of associated disorders, as targets in microarray, to generate hybridization probes, and to detect differences in gene sequences among normal, carrier or affected individuals. Antibodies may also be used in diagnosing disorders, in monitoring patients being treated with EXCS agonists, antagonists or inhibitors. Sequences c84293-c84318 represent nucleic acid molecules encoding the EXCS of the invention.

Sequence 1055 BP; 284 A; 254 C; 278 G; 239 T; 0 other;

Query Match 36.6%; Score 439.8; DB 22; Length 1055;

|                       |        |                     |        |               |
|-----------------------|--------|---------------------|--------|---------------|
| Query Match           | 50.0%, | Score 433.0,        | DB 22, | Bengali 1055, |
| Best Local Similarity | 77.48; | Pred. No. 1.9e-123; |        |               |

| Best Local Similarity     | 77.4% | Filed; NO: 1.9e | 125;               |
|---------------------------|-------|-----------------|--------------------|
| Matches 600; Conservative | 0;    | Mismatches 162; | Indels 13; Gaps 5; |

|    |     |  |     |
|----|-----|--|-----|
| Qy | 18  | ACAGATTAGACTCAGAACCAAGAGGTCGATAGGGGTCCA-CAGGGCTGCTGCAGAT         | 76  |
| Db | 75  | agaaatcagctcaagaagtggaagtggcagagattccaccaggaactggtgcaga          | 134 |
| Qy | 77  | CACAGAGCCAGCCAGATCTCAGAGAGCAGGCAACAAGATGCTGGATTCGACAGAGCAATAA    | 136 |
| Db | 135 | cgagagccagcagaattgagaagaaggaacaaagatgctggggagcagagctgtaatg       | 194 |
| Qy | 137 | ATGCTATGGCTGTGGCTGGGTCACTCAGGGCTGGCTGTGCTTAGGAGTAGCACTCCT        | 196 |
| Db | 195 | ctgctgttctgtgcctggacagctcagggcagagactgctcgtgggcagcagccct         | 254 |
| Qy | 197 | GACTGGGCTCAGTGCCAGCAGCTCTCTCGGAATCTCTGCATGCTAGCTCGGAACGCACAT     | 256 |
| Db | 255 | gcctggactcaatgccagaagctttcaagaagctctgcacactggcctggagtgcaact      | 314 |
| Qy | 257 | GCACAGGGGACATATGAATCTACTTAAGAGAGAAGAGAGGATGAAGAGACTATAAATAAT     | 316 |
| Db | 315 | ccactagtgggacacatggat---ctaagagagagaggatgaagagactacaaatgat       | 371 |
| Qy | 317 | GTGCCCCGTATPCCAGTGTGAAGATGPTGTGACCCCAAGGACTCAAGAGACAACACCCAG     | 376 |
| Db | 372 | gtcccccatatccagtgtgagatgycctgtgacccccaaaggaactcaggggcaaacagtcaag | 431 |
| Qy | 377 | TTCTGCTGTGCNAAGGATCCGCCAAGGPTCTGGCTTTTATTAAGCACCTGCTTGACTCTGAC   | 436 |
| Db | 432 | tctcgtgtcaaaagatccaccagggtcgtgatttttatgagaagctgctgaggtatcggt     | 491 |
| Qy | 437 | ATCTTTAAAGGGGAGCTGCTCTACTCCCTGATAGCCCCCATGAGCAACTTCACACCTCC      | 496 |
| Db | 492 | attttcacaggggagcctctctcgtctccccctgaagccccctggggcagcttcagctcc     | 551 |
| Qy | 497 | CTACTAGGACTCAGCCAACTCCTCCAGCCAGAGGATATACCCCCGGGAGACCCCAACAGATG   | 556 |
| Db | 552 | ctactgggctcagcccaactcctgcagccttgaaaggtccaccacttggaagactcagcaatt  | 611 |





















GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2001, 01:24:55 ; Search time 1723.79 Seconds  
(without alignments)  
6596.965 Million cell updates/sec

Title: US-09-558-474-3  
Perfect score: 1203  
Sequence: 1 CGCTTAGAAGTCGACTACA.....TACCANTACGAAGTACAAA 1203

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

|     |              |
|-----|--------------|
| 1:  | gb_est1:*    |
| 2:  | gb_est2:*    |
| 3:  | gb_est3:*    |
| 4:  | gb_est4:*    |
| 5:  | gb_est5:*    |
| 6:  | gb_est6:*    |
| 7:  | gb_est7:*    |
| 8:  | gb_est8:*    |
| 9:  | gb_est9:*    |
| 10: | gb_est10:*   |
| 11: | gb_est11:*   |
| 12: | gb_est12:*   |
| 13: | gb_est13:*   |
| 14: | gb_est14:*   |
| 15: | gb_est15:*   |
| 16: | gb_est16:*   |
| 17: | gb_est17:*   |
| 18: | gb_est18:*   |
| 19: | gb_est19:*   |
| 20: | gb_est20:*   |
| 21: | gb_est21:*   |
| 22: | gb_est22:*   |
| 23: | gb_est23:*   |
| 24: | gb_est24:*   |
| 25: | gb_est33:*   |
| 26: | gb_est34:*   |
| 27: | gb_est35:*   |
| 28: | gb_est36:*   |
| 29: | gb_est37:*   |
| 30: | gb_est38:*   |
| 31: | gb_est39:*   |
| 32: | gb_est40:*   |
| 33: | em_estba:*   |
| 34: | em_estfun:*  |
| 35: | em_esthum1:* |
| 36: | em_esthum2:* |
| 37: | em_esthum3:* |
| 38: | em_esthum4:* |
| 39: | em_esthum5:* |
| 40: | em_esthum6:* |
| 41: | em_esthum7:* |
| 42: | em_esthum8:* |
| 43: | em_esthum9:* |

|      |               |
|------|---------------|
| 44:  | em_esthum10:* |
| 45:  | em_esthum11:* |
| 46:  | em_esthum12:* |
| 47:  | em_esthum13:* |
| 48:  | em_esthum14:* |
| 49:  | em_esthum15:* |
| 50:  | em_esthum16:* |
| 51:  | em_esthum17:* |
| 52:  | em_esthum18:* |
| 53:  | em_esthum19:* |
| 54:  | em_esthum20:* |
| 55:  | em_esthum21:* |
| 56:  | em_esthum22:* |
| 57:  | em_esthum23:* |
| 58:  | em_esthum24:* |
| 59:  | em_esthum25:* |
| 60:  | em_esthum26:* |
| 61:  | em_esthum27:* |
| 62:  | em_esthum28:* |
| 63:  | em_estin1:*   |
| 64:  | em_estin2:*   |
| 65:  | em_estin3:*   |
| 66:  | em_estin4:*   |
| 67:  | em_estin5:*   |
| 68:  | em_estom1:*   |
| 69:  | em_estom2:*   |
| 70:  | em_estov1:*   |
| 71:  | em_estov2:*   |
| 72:  | em_estpl1:*   |
| 73:  | em_estpl2:*   |
| 74:  | em_estpl3:*   |
| 75:  | em_estpl4:*   |
| 76:  | em_estpl5:*   |
| 77:  | em_estpl6:*   |
| 78:  | em_estpl7:*   |
| 79:  | em_estpl8:*   |
| 80:  | em_estpl9:*   |
| 81:  | em_estpl10:*  |
| 82:  | em_estro1:*   |
| 83:  | em_estro2:*   |
| 84:  | em_estro3:*   |
| 85:  | em_estro4:*   |
| 86:  | em_estro5:*   |
| 87:  | em_estro6:*   |
| 88:  | em_estro7:*   |
| 89:  | em_estro8:*   |
| 90:  | em_estro9:*   |
| 91:  | em_estro10:*  |
| 92:  | em_estro11:*  |
| 93:  | em_estro12:*  |
| 94:  | em_estro13:*  |
| 95:  | em_estro14:*  |
| 96:  | em_estro15:*  |
| 97:  | em_estro16:*  |
| 98:  | em_estro17:*  |
| 99:  | em_estro18:*  |
| 100: | em_estro19:*  |
| 101: | em_estro20:*  |
| 102: | gb_est25:*    |
| 103: | gb_est26:*    |
| 104: | gb_est27:*    |
| 105: | gb_est28:*    |
| 106: | gb_est29:*    |
| 107: | gb_est30:*    |
| 108: | gb_est31:*    |
| 109: | gb_est32:*    |
| 110: | gb_est41:*    |
| 111: | gb_est42:*    |
| 112: | gb_est43:*    |
| 113: | gb_est44:*    |
| 114: | gb_est45:*    |
| 115: | gb_est46:*    |
| 116: | gb_est47:*    |

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                 |
|------------|-------|-------------|--------|-------|-----------------------------|
| 1          | 638   | 53.0        | 1150   | 137   | BE572471 601331259          |
| 2          | 481.4 | 40.0        | 484    | 141   | BE850725 ux01401.y          |
| 3          | 479   | 39.8        | 897    | 137   | BE571481 601333907          |
| 4          | 429   | 35.7        | 430    | 141   | BE847383 uw26b04.y          |
| 5          | 384   | 31.9        | 484    | 166   | BE335675 us87h10.y          |
| 6          | 368   | 30.6        | 523    | 147   | BF318114 ux01401.x          |
| 7          | 340   | 28.3        | 469    | 141   | BE851424 ux05h11.y          |
| 8          | 299.8 | 24.9        | 833    | 104   | AJ277049 AJ277049           |
| 9          | 298   | 24.8        | 323    | 129   | BB270313 BB270313           |
| 10         | 295.2 | 24.5        | 320    | 130   | BB274745 BB274745           |
| 11         | 295   | 24.5        | 458    | 143   | BF018105 BF018105           |
| 12         | 290.8 | 24.2        | 492    | 7     | AA418955 zw01c10.r          |
| 13         | 281   | 23.4        | 447    | 123   | AW987117 uf14d07.x          |
| 14         | 183.8 | 15.3        | 372    | 155   | C06368 C06368 Huma          |
| 15         | 172.4 | 14.3        | 542    | 102   | AI796983 w25e07.x           |
| 16         | 151.8 | 12.6        | 719    | 153   | BG435629 602506911          |
| 17         | 135.4 | 11.3        | 496    | 175   | BG236314 naf26b02.          |
| 18         | 121.6 | 10.1        | 575    | 31    | AV598349 AV598349           |
| 19         | 120   | 10.0        | 367    | 145   | BF189762 235459 MA          |
| 20         | 109.6 | 9.1         | 399    | 114   | AW316433 14841 MAR          |
| 21         | 108.8 | 9.0         | 457    | 117   | AW511070 h038e10.x          |
| 22         | 96.8  | 8.0         | 566    | 165   | BE237060 146167 MA          |
| 23         | 78.8  | 6.6         | 137    | 188   | TI8577 hbc2125 Hum          |
| 24         | 71.2  | 5.9         | 400    | 14    | AI004049 oul1d12.x          |
| 25         | 46    | 3.8         | 1101   | 219   | CNS01784 AL107710 Drosophil |
| 26         | 45.8  | 3.8         | 997    | 219   | CNS0057E AL060767 Drosophil |
| 27         | 45.6  | 3.8         | 369    | 111   | AW072323 w29h07.x           |
| 28         | 45.4  | 3.8         | 943    | 219   | CNS002IM AL077720 Drosophil |
| 29         | 44.8  | 3.7         | 987    | 219   | CNS00418 AL066537 Drosophil |
| 30         | 44.8  | 3.7         | 990    | 219   | CNS017DC AL107898 Drosophil |
| 31         | 43.4  | 3.6         | 842    | 256   | B19240 F9C2-T7 IGF          |
| 32         | 42.8  | 3.6         | 397    | 7     | AA418747 zw01c10.s          |
| 33         | 42.6  | 3.5         | 1101   | 219   | CNS017GS AL108022 Drosophil |
| 34         | 42.4  | 3.5         | 1101   | 219   | CNS0181N AL108773 Drosophil |
| 35         | 42    | 3.5         | 1101   | 219   | CNS017RP AL108415 Drosophil |
| 36         | 41.8  | 3.5         | 698    | 32    | AV728560 AV728560           |
| 37         | 40.6  | 3.4         | 305    | 229   | AQ512257 HS 5081.A          |
| 38         | 40.4  | 3.4         | 957    | 105   | AL526604 AL526604           |
| 39         | 40.2  | 3.3         | 937    | 142   | BE964027 60165796           |
| 40         | 40.2  | 3.3         | 1100   | 219   | CNS016KD AL106855 Drosophil |
| 41         | 40    | 3.3         | 2092   | 192   | AK018120 Mus muscu          |
| 42         | 39.8  | 3.3         | 799    | 80    | BF262498 HV_CEA000          |
| 43         | 39.8  | 3.3         | 1159   | 219   | CNS015XR AL106041 Drosophil |
| 44         | 39.6  | 3.3         | 633    | 256   | B19596 T30K12-T7.1          |
| 45         | 39.6  | 3.3         | 1101   | 219   | CNS006BP AL064052 Drosophil |

ALIGNMENTS

| RESULT | LOCUS    | DEFINITION                | Accession | Version | Keywords   | Source      | Organism     | Reference           | Authors                           | Title                   | Journal  | Comment            |
|--------|----------|---------------------------|-----------|---------|------------|-------------|--------------|---------------------|-----------------------------------|-------------------------|--|--------------------|
| 1      | BE572471 | 601331259F1 NCI_CGAP_Mam6 | BE572471  | 1       | GI:9816191 | house mouse | Mus musculus | 1 (bases 1 to 1150) | NIH-MGC                           | http://mgc.nci.nih.gov/ | National Institutes of Health, Mammalian Gene Collection (MGC) | Unpublished (1999) |
|        |          |                           |           |         |            |             |              |                     | Contact: Robert Strausberg, Ph.D. |                         |  |                    |

Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM8928 row: f column: 16  
High quality sequence stop: 716.  
Location/Qualifiers  
1. 1150  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3708687"  
/clone\_lib="NCI\_CGAP\_Mam6"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"  
BASE COUNT 330 a 259 c 329 g 231 t 1 others  
ORIGIN

Query Match 53.0%; Score 638; DB 137; Length 1150;  
Best Local Similarity 89.7%; Pred. No. 1.8e-174;  
Matches 788; Conservative 0; Mismatches 75; Indels 15; Gaps 9;  
QY 37 CAAGAGAGTGTGATGGGGTCCACAGGCTGTGGTGTGATGCTGAGTGGGTCAGTGC 95  
Db 1 CAAGAGAGTGTGATGGGGTCCACAGGCTGTGGTGTGATGCTGAGTGGGTCAGTGC 60  
QY 96 GAGAGCAGGGAACAGAGTGTGGATGGCAGACAGTAAATGCTA-TGGCTGTGGCC 154  
Db 61 GAGAGCAGGGAACAGAGTGTGGATGGCAGACAGTAAATGCTGAGTGGGTCAGTGC 120  
QY 155 TGGGTCACTCAGGGCC-TGGCTGTGCC-TAGGAGTAGCAGTCTGACTGGGTCAGTGC 212  
Db 121 TGGGTCACTCAGGGCCGTGGCTGTGGCTGTGGAGTAGCAGTCTGACTGGGTCAGTGC 180  
QY 213 AGCAGCTCTCTCGAATCTCTGC--ATGCTAGCTGTGAACGACATGCACAGCGGACA 270  
Db 181 AGCAGCTCTCTCGAATCTCTGCAGTGTGGAGTAGCAGTCTGACTGGGTCAGTGC 240  
QY 271 TATGAATCTTAAAGAGAAGAGGATCAAGAGACTAAATAATATGTCCTCCGATCCA 330  
Db 241 TATGAATCTTAAAGAGAAGAGGATCAAGAGACTAAATAATATGTCCTCCGATCCA 300  
QY 331 GTGTGAAGATGGTGTGTGACCCCAAGGACTCAAGGACAACAGCAGTCTTGTGTCAGAG 390  
Db 301 GTGTGAAGATGGTGTGTGACCCCAAGGACTCAAGGACAACAGCAGTCTTGTGTCAGAG 360  
QY 391 GATCGCCCAAGTCTGGCTTTTATAAGCACTGCTTGTGACTGTGACATCTTCAAGGGGA 450  
Db 361 GATCGCCCAAGTCTGGCTTTTATAAGCACTGCTTGTGACTGTGACATCTTCAAGGGGA 420  
QY 451 GCCTGCTCTACTCCC-TGATAGCCCCCATGAGCACTTCCACACCTCCCTACTAGGACTCA 509  
Db 421 GCCTGCTCTACTCCC-TGATAGCCCCCATGAGCACTTCCACACCTCCCTACTAGGACTCA 480  
QY 510 GCCAATCTCTCAGGAGGATCACCCCGGGAGACCCCAACAGATGCCAGCC-TGAGT 568  
Db 481 GCCAATCTCTCAGGAGGATCACCCCGGGAGACCCCAACAGATGCCAGCC-TGAGT 540  
QY 569 TCTAGTACAGTGGCAGCGCCCTTCTCCGTTTCAAGATCTTCAAGATCTTCAAGATCTTCAAG 628  
Db 541 TCTAGTACAGTGGCAGCGCCCTTCTCCGTTTCAAGATCTTCAAGATCTTCAAGATCTTCAAG 600

```

QY 629 TTTTGGCCATAGCTGCCCGGGTCTTTGCCACGAGCAGCAACTCTGACTGAGCCCTTA 688
|||||
Db 601 TTTTGGCCATAGCTGCCCGGGTCTTTGGCCACGAGCAGCAACTCTGACTGAGCCCTTA 660
QY 689 GTGCCAACAGCTTAAGCATGCCCGAGGTCCCATGGCTTACCATGATAGACTAATCTATCA 748
|||||
Db 661 GTGCAACAGCTTAAGGATGCCAAGGTTCCTCATGGTACCATGATAGATAATCTATCAG 720
QY 749 GCCCAGACATCTACCACTTAATTAACCCATTAGGACTTGTGCTTCTTCTTCTTGT 808
|||||
Db 721 GCCAAGACATCTACCACTTAAGTATTCATTAGGATTGTGGGAACTGGAAGGTGGAGCG - 779
QY 809 TTTGCGTGAAGGCAAGCAGCACCATATTATAAGAGAGAAAAGAACCAACCCAGAGCA - GG 867
|||||
Db 780 ----GTGAAGGCGAGGACCATTTATCAAGAGAAAGTAAACCAACCCAGGAGCAGG 834
QY 868 CAGCTGCTAGAGAAAGAGCTGGAGAGAGAGATAAA 905
|||||
Db 835 CAGCGGTAACACAAAGGAGACTGTGTCAAGAGAGAAA 872

RESULT 2
BE850725 484 bp mRNA EST 26-SEP-2000
LOCUS ux01d01.y1 Soares_thymus_2NBMT Mus musculus cDNA clone
DEFINITION IMAGE:3470209 5', mRNA sequence.
ACCESSION BE850725
VERSION BE850725
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1389569
Seq primer: -40RP from Gibco
High quality sequence stop: 459.
FEATURES
source
1. .484
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3470209"
/sex="male"
/tissue_type="thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCATCTGAAGTGGAGCGCGCGGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 144 a 109 c 130 g 100 t 1 others
ORIGIN

Query Match 40.0%; Score 481.4; DB 141; Length 484;
Best Local Similarity 99.6%; Pred. No. 5.3e-129;
Matches 482; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 658 CCAGGAGCAGCAACTCTGACTGAGCCCTTAGTGCCAAACAGCTTAAGGATGCCAGGTTT 717
|||||
Db 1 CCAGGAGCAGCAACTCTGACTGAGCCCTTAGTGCCAAACAGCTTAAGGATGCCAGGTTT 60
QY 718 CCATGGCTACCATGATAAGACTAATCTATCAGCCCAAGACATCTACCACTTAATTAACCCA 777
|||||
Db 61 CCATGGCTACCATGATAAGACTAATCTATCAGCCCAAGACATCTACCACTTAATTAACCCA 120
QY 778 TTAGGACTTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 837
|||||
Db 121 TTAGGACTTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 180
QY 838 AAAGAGAAAAAGAAACAAACCCAGAGCAGCAGCTGCTAGAGAAAGAGCTGGAGAGA 897
|||||
Db 181 AAAGAGAAAAAGAAACAAACCCAGAGCAGCAGCTGCTAGAGAAAGAGCTGGAGAGA 240
QY 898 AGAATAAAGTCTCGAGCCCTTGGCCTTGGAGGGGCAAGCAGCTGCGTGGCCCTGAGGGG 957
|||||
Db 241 AGAATAAAGTCTCGAGCCCTTGGCCTTGGAGGGGCAAGCAGCTGCGTGGCCCTGAGGGG 300
QY 958 AGGGGGCGTGGCATCGAGAAACTGTGAGAAAAACCCAGAGCATCAGAAAAAGTGAGCCC 1017
|||||
Db 301 AGGGGGCGTGGCATCGAGAAACTGTGAGAAAAACCCAGAGCATCAGAAAAAGTGAGCCC 360
QY 1018 AGCTTTGGCCATTATCTGTAAAGAAAAAAGAAAGGGGAACATTTACTTTCTCTGGT 1077
|||||
Db 361 AGCTTTGGCCATTATCTGTAAAGAAAAAAGAAAGGGGAACATTTACTTTCTCTGGT 420
QY 1078 GGCTCAGGGAATGTGCAGATGCACAGTACTTCCAGAGCAGCTGTGTACCTGCTCTC 1137
|||||
Db 421 GGCTCAGGGAATGTGCAGATGCACAGTACTTCCAGAGCAGCTGTGTACCTGCTCTC 480
QY 1138 TGTC 1141
|||||
Db 481 TGTC 484

RESULT 3
BE571481 897 bp mRNA EST 15-AUG-2000
LOCUS 601333907F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3711529 5',
DEFINITION mRNA sequence.
ACCESSION BE571481
VERSION BE571481
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8935 row: m column: 02
High quality sequence stop: 597.
FEATURES
source
1. .897
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3711529"
/sex="female, virgin"

```

```
/tissue_type="infiltrating ductal carcinoma"  
/dev_stage="5 months"  
/lab_host="nu10p"
```

```

/label=Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
201 c 262 g 194 t 1 others.

```

Query Match 39.8%; Score 479; DB 137; Length 897;  
Best Local Similarity 93.5%; Pred. No. 3.2e-128;  
Matches 609; Conservative 0; Mismatches 30; Indels 12; Gaps 10;

|            |   |             |
|------------|---|-------------|
| RESULT     | 4   |             |
| BE847383   |   |             |
| LOCUS      | BE847383  |             |
| DEFINITION | 430 bp mRNA   | 26-SEP-2000 |
|            | uw26b04.y1 Soares_thymus_2NbMT Mus musculus cDNA clone    |             |
|            | IMAGE:3417775 5', mRNA sequence.                          |             |
| ACCESSION  | BE847383  |             |
| VERSION    | BE847383.1  | GI:10305722 |
| KEYWORDS   | EST.  |             |
| SOURCE     | house mouse.  |             |
| ORGANISM   | Mus musculus  |             |
|            | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;   |             |
|            | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; |             |
| REFERENCE  | 1 (bases 1 to 430)  |             |



```

QY 1076 GTGGCTCAGGAAATGTCAGATGCACAGTACTCCAGACAGCAGCTCTCTAGCTGCCTGC 1135
|||||
Db 283 GTGGCTCAGGAAATGTCAGATGCACAGTACTCCAGACAGCAGCTCTCTAGCTGCCTGC 224
QY 1136 TCTGTCCCTCAGTCTTAACAGAAATCTAGTCTACTAGAACTACAGACTACCAATACGAA 1195
|||||
Db 223 TCTGTCCCTCAGTCTTAACAGAAATCTAGTCTACTAGAACTACAGACTACCAATACGAA 164
QY 1196 CTGACAAA 1203
|||||
Db 163 CTGACAAA 156

RESULT 7
BE851424
LOCUS BE851424 469 bp mRNA EST 26-SEP-2000
DEFINITION uw95h11.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:3469797 5', mRNA sequence.
ACCESSION BE851424
VERSION BE851424.1 GI:10309763
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 469)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1389157
Seq primer: -40RP from Gibco
High quality sequence stop: 396.

FEATURES
source
1. 469
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3469797"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 147 a 89 c 118 g 115 t
ORIGIN

Query Match 28.3%; Score 340; DB 141; Length 469;
Best Local Similarity 100.0%; Pred. No. 7.7e-88;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 864 CAGGCAGCTGGCTAGAGAAAGAGCTGGAGAGAGAGATAAAGTCTCGAGCCCTTGGCCCT 923
|||||
Db 1 CAGGCAGCTGGCTAGAGAAAGAGCTGGAGAGAGAGATAAAGTCTCGAGCCCTTGGCCCT 60
QY 924 TGAAGCGGGCAAGCAGCTGCGCTCGAGGAGGGGGCGGTGGCATCGAGAACTG 983
|||||
bb 61 TGAAGCGGGCAAGCAGCTGCGCTCGAGGAGGGGGCGGTGGCATCGAGAACTG 120
QY 984 TGAGAAAACCCAGAGCATCAGAAAAGAGTGAGCCCGCTTTGGCCCATTTATCTGTAGAAA 1043
|||||
Db 121 TGAGAAAACCCAGAGCATCAGAAAAGAGTGAGCCCGCTTTGGCCCATTTATCTGTAGAAA 180
|||||

```

```

QY 1044 AACAAGAAAGGGAGCAATTATCTCTGGTGGCTCAGGAAATGTGCAGATGCCACA 1103
|||||
Db 181 AACAAGAAAGGGAGCAATTATCTCTGGTGGCTCAGGAAATGTGCAGATGCCACA 240
QY 1104 GTACTCCAGACAGCAGCTCTGTACTCTGCTCTGTCTCTCTCAGTTCTTAACAGAATCTAG 1163
|||||
Db 241 GTACTCCAGACAGCAGCTCTGTACTCTGCTCTGTCTCTCAGTTCTTAACAGAATCTAG 300
QY 1164 TCACCTAAGAACTAACAGGAGTACCAATAGCAACTGACAAA 1203
|||||
Db 301 TCACCTAAGAACTAACAGGAGTACCAATAGCAACTGACAAA 340

RESULT 8
AJ277049
LOCUS AJ277049 833 bp mRNA EST 30-MAR-2000
DEFINITION AJ277049 Placenta Matchmaker cDNA library(Clontech) Homo sapiens
cDNA clone BH42, mRNA sequence.
ACCESSION AJ277049
VERSION AJ277049.1 GI:7342435
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 833)
AUTHORS Bach,S., Portetelle,D. and Vandenbol,M.
TITLE Isolation of new human EST sequences by using the MSB4 yeast gene
as bait in the two-hybrid screening of a Human Placenta cDNA
library
JOURNAL Unpublished (2000)
COMMENT Contact: Bach SL
Microbiology Unit
Gembloux Agricultural University
Avenue Marechal Juin 6, B-5030 Gembloux, Belgium.

FEATURES
source
1. 833
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BH42"
/clone_lib="Placenta Matchmaker cDNA library(Clontech)"
/sex="Female"
/tissue_type="Placenta"
BASE COUNT 227 a 197 c 200 g 209 t
ORIGIN

Query Match 24.9%; Score 299.8; DB 104; Length 833;
Best Local Similarity 78.3%; Pred. No. 4.8e-76;
Matches 400; Conservative 0; Mismatches 102; Indels 9; Gaps 3;

QY 281 CTAAGAGAAAGAGGATGAAGAGACTATAAATAATGTGCCCGTATCCAGTGTGAAGAT 340
|||||
Db 100 CTAAGAGAAAGAGGATGAAGAGACTATAAATAATGTGCCCGTATCCAGTGTGAAGAT 159
QY 341 GGTGTGACCCACAGGAGTCAAGGACACAGCCAGTCTGTGTTGCAAGAGTCCGCCAA 400
|||||
Db 160 GGCTGTGACCCACAGGAGTCAAGGACACAGTCAAGTCTGTGTTGCAAGAGTCCACCAG 219
QY 401 GGTGTGCGCTTTTATAGACACCTGCTTGACTCTGACATCTTCAAGGGGAGCGCTGCTTA 460
|||||
Db 220 GGTCTGATTTTATAGAAAGCTGTAGGATCGGATATTTTACAGGGGAGCGCTTCTCTG 279
QY 461 CTCCTCTGATAGCCCCATGGAGCAACTTTCACACTCCCTACTAGGACTCAGCCCAACTCTC 520
|||||
Db 280 CTCCTCTGATAGCCCCATGGAGCAACTTTCACACTCCCTACTAGGACTCAGCCCAACTCTC 339
QY 521 CAGCCAGAGGATACACCCCGGAGAGACCCCAACAGATGCCAGCGCTGTAGTTCTAGTCAGCAG 580
|||||
Db 340 CAGCGCTGAGGGTCAACACTGGGAGAGTCAAGAGCTCAGTCAAGCGCTCAGTCCAGCCAGCCA 399
QY 581 TGGCAGCGCCCCCTTCTCCGTTCCAGAGATCCTTCGAGAGCCTCCAGGCGCTTTTGGGCCATA 640
|||||

```







```
Qy 969 GGCATCGAGAACTGTGAGAAAACCCAGAGCATCAGAAAAAGTGGCCAGGCTTTGGCC 1028
|||||
Db 398 GGCATCGAGAACTGTGAGAAAACCCAGAGCATCAGAAAAAGTGGCCAGGCTTTGGCC 339

Qy 1029 ATTATCTGTAGAAAAACAAAGAAAGGGGACATATTATCTTTCTCGGGTGCTCAGGGAA 1088
|||||
Db 338 ATTATCTGTAGAAAAACAAAGAAAGGGGACATATTATCTTTCTCGGGTGCTCAGGGAA 279

Qy 1089 ATGTGCAGATCGACAGTACTCCACAGACGAGCTGTGTACCTGCCTGCTCTGCTCCTCAGT 1148
|||||
Db 278 ATGTGCAGATCGACAGTACTCCACAGACGAGCTGTGTACCTGCCTGCTCTGCTCCTCAGT 219

Qy 1149 TCTAACGAGTCTAGTCTACTAAGAACTAACAGGAGTACCTACCAATAGAACTGACAAA 1203
|||||
Db 218 TCTAACGAGTCTAGTCTACTAAGAACTAACAGGAGTACCTACCAATAGAACTGACAAA 164

RESULT 12
AA418955
LOCUS AA418955 492 bp mRNA EST 16-OCT-1997
DEFINITION zw01c10.r1 Soares_NhMpu_S1 Homo sapiens cDNA clone IMAGE:768018
5' mRNA sequence.
ACCESSION AA418955
VERSION AA418955.1 GI:2080765
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 492)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisler,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie
T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham.

FEATURES
source
Location/Qualifiers
1..492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_NhMpu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below); Vector: p7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NhMpu, and fetal heart NhH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT 117 a 136 c 136 g 103 t
ORIGIN

Query Match 24.28; Score 290.8; DB 7; Length 492;
Best Local Similarity 84.18; Pred. No. 1.7e-73;
Matches 328; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

```
Qy 281 CTAAGAGAAGAGGAGTGAAGAGACTAAAATAATGTGCCCGCTATCCAGTGTGAAGAT 340
|||||
Db 103 CTAAGAGAAGAGGAGTGAAGAGACTACAATAATGTGCCCGCTATCCAGTGTGAAGAT 162

Qy 341 GGTGTGTGACCCACAAAGAGCTCAAGGACAACAGCCAGTTCGTCTGTGCAAAAGATCCGCCAA 400
|||||
Db 163 GGTGTGTGACCCACAAAGAGCTCAAGGACAACAGTTCGTCTGTGCAAAAGATCCGCCAG 222

Qy 401 GGTCTGCTTTTATATAGCACCTGCTTGTGACTCTCTGACATCTTCAAAGGGAGGCTGTCTTA 460
|||||
Db 223 GGTCTGATTTTATGAGAAGCTCTAGGATCGGATATTTTACAGAGGGAGGCTTCTCTG 282

Qy 461 CTCCTGTATAGCCCATCGGAGCAACTTTCACACCTCCCTACTAGGACTCAGCAACTCCTC 520
|||||
Db 283 CTCCTGTATAGCCCTGTGGCGAGCTTCATGCTCCCTACTGTGGCCTCAGCAACTCCTG 342

Qy 521 CAGCCAGAGTACACCCCGGGAGACCAACAGATGCCAGCCTGTAGTTCAGTTCAGCAG 580
|||||
Db 343 CAGCCTGAGGCTCACCACTGGGAGACTCAGCAGATTCGAAGCCTCAGTCCAGCAGCCA 402

Qy 581 TGGCAGCGCCCTTCTCCGTTCCAAGATCCTTCGAAGCCTCCAGGCTTTTGTGCCATA 640
|||||
Db 403 TGGCAGCGCTCTCTCTCCGCTTCAAAATCTTCGACGCTCCAGGCTTTGTGGCTGTA 462

Qy 641 GCTGCCCGGGTCTTTGGCCAGGAGCAGCA 670
|||||
Db 463 GCCGCCCGGGTCTTTGCCAGGAGCAGCA 492

RESULT 13
AW987117/c
LOCUS AW987117 447 bp mRNA EST 02-JUN-2000
DEFINITION uf14d07.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:1511341 3', mRNA sequence.
ACCESSION AW987117
VERSION AW987117.1 GI:8181925
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 447)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:938193.

FEATURES
source
Location/Qualifiers
1..447
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 120 a 102 c 79 g 146 t
ORIGIN
```

Query Match 23.4%; Score 281; DB 123; Length 447;

Best Local Similarity 100.0%; Pred. No. 1.2e-70;

Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 923 TTGGGAAGCGGCAAGCAGCTGGCTGGCTGAGGGAAGGGCGGTGGCATCGAGAAACT 982
DB 447 TTGGAAGCGGGAAGCAGCTGGCTGGCTGAGGGAAGGGCGGTGGCATCGAGAAACT 388
QY 983 GTGAGAAACCCAGAGCATCAAAAAGTGAGCCAGGCTTTGGCCATTATCTGTAAGAA 1042
DB 387 GTGAGAAACCCAGAGCATCAAAAAGTGAGCCAGGCTTTGGCCATTATCTGTAAGAA 328
QY 1043 AAACAAGAAAGGGAACATTATCTTCTCGGGTGGCTCAGGGAATGTGACAGATGCAC 1102
DB 327 AAACAAGAAAGGGAACATTATCTTCTCGGGTGGCTCAGGGAATGTGACAGATGCAC 268
QY 1103 AGTACTCCAGACAGCAGCTCTGTACTGCTGCTCTGTCCCTCAGTCTTAACAGAACTCTA 1162
DB 267 AGTACTCCAGACAGCAGCTCTGTACTGCTGCTCTGTCCCTCAGTCTTAACAGAACTCTA 208
QY 1163 GTCACTAAGAACTAACAGGACTACCAATACGAACCTGACAAA 1203
DB 207 GTCACTAAGAACTAACAGGACTACCAATACGAACCTGACAAA 167
```

## RESULT 14

LOCUS C06368 372 bp mRNA EST 16-OCT-1996  
DEFINITION C06368 Human pancreatic islet Homo sapiens cDNA clone hbc5699, mRNA sequence.

ACCESSION C06368

VERSION C06368.1 GI:1503144

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Takeda, J.

TITLE Human pancreatic islet ESTs

JOURNAL Unpublished (1995)

CONTACT: Jun Takeda

Institute for Molecular and Cellular Regulation, Gunma University 3-39-15 Showa-machi, Maebashi Gunma 371, Japan

Tel: 272-20-8856

Fax: 272-20-8896

Email: jtakeda@sb.gunma-u.ac.jp.

Location/Qualifiers

1..372

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="hbc5699"

/clone\_lib="Human pancreatic islet"

/note="Vector: Lambda ZAPII; Site\_1: Eco RI; Site\_2: Xho I; mRNA was prepared from normal adult human islets. cDNA was directionally synthesized from the Xho I in the vector to the EcoRI site. cDNA was size fractionated to remove sequences <1000 bp in size."

BASE COUNT 107 a 88 c 116 g 59 t 2 others

ORIGIN

## Query Match

Best Local Similarity 15.3%; Score 183.8; DB 155; Length 372;

Matches 248; Conservative 0; Mismatches 69; Indels 4; Gaps 2;

```
QY 18 ACAGAGTTAGACTCAGAACCAAGAGGTGGATGGGGTCCA-CAGGCGCTGGTGACAGAT 76
DB 55 AGAGATCAGGCTCAAGCAAGTGGAGTGGCAGAGATTCCACCAGGACTGGTGCAGG 114
```

```
QY 77 CACAGAGCCAGCCAGTCTGAGACAGGGAACAGATGCTGGATTGCAGACAGTAATA 136
DB 115 CGCAGAGCCAGCCAGTCTGAGACAGGGAAGCAAAAGATGCTGGGAGCAGAGCTGTAATG 174
```

```
QY 137 ATGCTATGGCTCTTGCCTGGGTCTACTCAGGCGCTGGCTGTGCCTAGGAGTAGCAGTCT 196
DB 175 CTGCTGTGTCTCTGCTGCGCTGGACAGCTCAGGGCAGAGCTGTGCTGGGGCAGCAGCC 234
QY 197 GACTGGGCTCAGTGGCAGCAGCTCTCTCGGAATCTCTGCATCTCTAGCCTAGCCTGAACGACAT 256
DB 235 GCCTGGACTCAGTGGCAGCAGCTTTTCACAGAAAGCTCTGCACACTGGCCTGGAGTGACAT 294
QY 257 GCACAGCGGGACATATGATCTACTAAGAGAAGAGAGGATGAAGAGATTAATAATAT 316
DB 295 COACTAGTGGGACATGATGAT---CTAAGAGAAGAGGAGATGAAGAGACTACNAATGAT 351
QY 317 GTGCCCGCTATCCAGTGTGAA 337
DB 352 GTTCCCATATCCAGTGTGGA 372
```

## RESULT 15

AI796983/c

LOCUS AI796983 542 bp mRNA EST 18-DEC-1999

DEFINITION we25e07.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:2342148 3', mRNA sequence.

ACCESSION AI796983

VERSION AI796983.1 GI:5362455

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 542)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapb-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 888 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 449.

Location/Qualifiers

1..542

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2342148"

/clone\_lib="NCI\_CGAP\_Lu24"

/tissue\_type="carcinoid"

/lab\_host="DH10B"

/note="organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI\_CGAP\_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 150 a 125 c 140 g 127 t

ORIGIN

## Query Match

Best Local Similarity 14.3%; Score 172.4; DB 102; Length 542;

Matches 257; Conservative 0; Mismatches 76; Indels 9; Gaps 3;

